

THESE

présentée devant

L'ECOLE NATIONALE SUPERIEURE AGRONOMIQUE DE RENNES

pour obtenir le titre de Docteur de l'ENSAR, spécialité Halieutique

par

François ROYER

Contribution à l'étude de la dynamique du thon rouge Atlantique

-

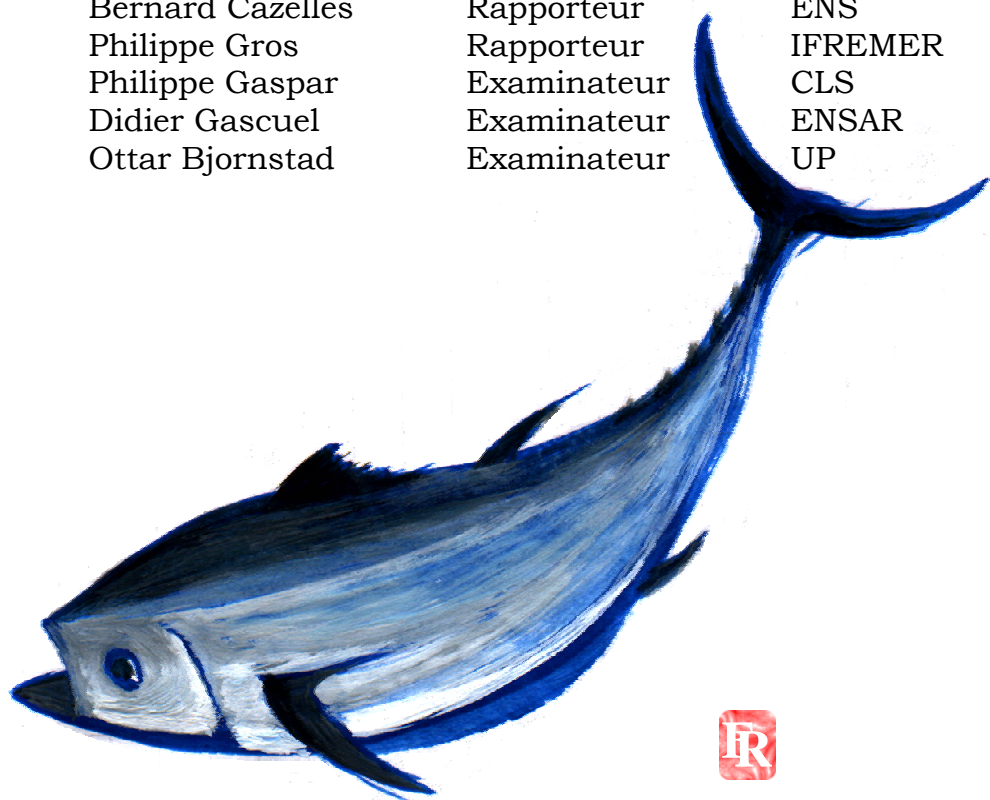
Approches écologiques et océanographiques

Préparée à la Direction Océanographie Spatiale de Collecte Localisation Satellites à Ramonville Saint-Agne, et au laboratoire Ifremer, Ressources Halieutiques du Centre de Recherches Halieutiques de Sète

Soutenue le

devant le jury composé de :

MM. Jean-Marc Fromentin	Directeur de thèse	IFREMER
Bernard Cazelles	Rapporteur	ENS
Philippe Gros	Rapporteur	IFREMER
Philippe Gaspar	Examinateur	CLS
Didier Gascuel	Examinateur	ENSAR
Ottar Bjornstad	Examinateur	UP



Remerciements

Aux nombreuses personnes qui ont contribué, de près ou de loin, ou même sans le savoir, à ces trois ans de réflexion.

A Jean-Marc Fromentin, dont j'ai eu la chance de croiser la route, pour ses qualités de guide et d'encadrant, pour sa connaissance des choses et sa rigueur, pour la richesse et la créativité de nos relations.

A Philippe Gaspar, pour la confiance qu'il m'a accordée, pour le cadre de travail unique au sein de CLS, et sa vision stimulante de l'océanobiologie,

A Bernard Cazelles et Philippe Gros, pour avoir accepté d'être les rapporteurs de cette thèse, tâche qui leur aura apporté, j'espère, autant de plaisir qu'à son auteur. A Didier Gascuel, président du jury et professeur, pour avoir grandement contribué à mon initiation à l'halieutique.

Aux habitants de CLS pour tous les moments partagés, en coup de vent d'autan ou de tramontane. Au bureau des thésards, Sido, Sabine, Isabelle, Annabelle pour avoir supporté mes frasques. A Philippe et Philippe pour leur aide précieuse en informatique hard et soft. A tous les voisins de bureau pour leur présence et leur patience quand "royer.bin" phagocytait CODA. A ceux d'en bas, Laurent et Olivier pour les délires "Got Fish?". Aux camarades de mer, Guillaume, Alix...

A la tribu Ifremer/IRD, après ma migration à Sete. A Franck et Manu pour tous les moments partagés. A Patrice "good job" pour ses conseils (du manuscrit à Bingin). A Eric pour le vent et les kilomètres de bitume. A Henri pour ses chimères. A Daniel pour son invraisemblance maximale. A Fred pour sa bonne humeur. A tous les habitants du couloir d'en haut et d'en bas (Olivier, Yunne, Francis, Pascal, Renaud...). A tous les intermittents de la recherche (Martin, Manu, Mathieu, Baptiste, Caro, Blaise, Remy, Tam, Audrey, Raphael...) pour le bout de chemin ensemble.

A mes frères.

A ma mère.

Durham, New Hampshire, le 3 juin 2005

Table des matières

AVANT-PROPOS	7
1. CONTEXTE OCEANOGRAPHIQUE ET ECOLOGIQUE	11
1.1. LE THON ROUGE ATLANTIQUE, <i>THUNNUS THYNNUS</i>	11
1.2. L'HABITAT OCEANIQUE	14
2. FRONTS OCEANQUES: UNE ECOLOGIE DES INTERFACES	21
2.1. ELEMENTS D'OCEANOGRAPHIE FRONTALE.....	21
2.2. LES FRONTS ET LA PRODUCTION BIOLOGIQUE.....	29
2.3. DE LA RELATION ENVIRONNEMENT-RESSOURCE	33
2.4. NECESSITE D'UNE APPROCHE DYNAMIQUE	44
3. DU MOUVEMENT DANS LE DOMAINE PELAGIQUE	47
3.1. REFORMULATION DE LA PROBLEMATIQUE.....	47
3.2. LA MARQUE ARCHIVE: TECHNOLOGIE ET PROMESSES	49
3.3. LE PROBLEME NUMERIQUE.	51
3.4. EVALUATION DE LA METHODE PARTICULAIRE	61
3.5. APPLICATION A UN CAS REEL	63
3.6. PERSPECTIVES IMMEDIATES	71
4. NOUVEAU REGARD SUR LES FLUCTUATIONS DE LA POPULATION DE <i>THUNNUS THYNNUS</i>.....	73
4.1. LA BASE DE DONNEES MADRAGUES	73
4.2. METHODES NON-LINEAIRES	ERROR! BOOKMARK NOT DEFINED.
4.3. RETOUR AUX TECHNIQUES CLASSIQUES	88
4.4. DISCUSSION SUR L'ANALYSE DES SERIES TEMPORELLES	104
5. SYNTHESE ET PERSPECTIVES.....	111
5.1. SYNTHESE SUR L'OBJET DE RECHERCHE, <i>THUNNUS THYNNUS</i>	111
5.2. PERSPECTIVES.....	116
BIBLIOGRAPHIE.....	121
ARTICLES ET MANUSCRITS.....	143

Avant-propos

L'écologie marine met à rude épreuve les facultés de raisonnement et les capacités calculatoires du scientifique. Parent pauvre du terrestre en termes d'observations et d'observabilité, appelant sans cesse à de nouveaux paradigmes, cette discipline a connu de nombreuses évolutions autour de son problème central, l'étude de la variabilité spatiale et temporelle des populations marines.

Reconnaissant la fragilité de l'écosystème océanique malgré ses grandes dimensions et son apparente inertie (Roberts and Hawkins 1999), la communauté halieutique a agrémenté sa vision d'un fonctionnement limité par les ressources (ou "bottom-up") d'une approche contrôlée par la prédation ou le prélèvement par pêche (ou "top-down"). La première approche a ainsi connu de nombreux succès dans la découverte de corrélations entre production primaire marine et abondance des niveaux trophiques supérieurs (Nixon 1988, Iverson 1990, Clark et al. 2001), *via* la conversion d'énergie solaire par photosynthèse. Toutefois la compréhension (et la prédiction) des fluctuations écologiques basée sur la seule connaissance des contrôles de cette productivité primaire peut s'avérer peu efficace (Hensen 1887, Verity et al. 2002). La non-linéarité et la dimension spatiale des processus incriminés en sont les principales causes: on observe ainsi des variations interannuelles d'un facteur 10 dans les captures ou le recrutement d'espèces de poissons pélagiques (Cushing 1982), contre des variations d'un facteur 2 seulement pour la production primaire associée (Lindahl and Hernroth 1988, Hubel et al. 1998). La prise en compte du forçage environnemental, déjà postulé par Hjort (1914), et de la nature stochastique des mécanismes écologiques a permis ainsi l'appréhension de nouveaux modes de variabilité (Cury and Roy 1989, Cushing 1990, Bjornstad et al. 1999, Stenseth et al. 2004). Aujourd'hui, l'interaction entre le changement climatique et la structure des écosystèmes est reconnue comme une source de variabilité majeure des populations marines (Skud 1982, Steele and Henderson 1984, McGowan et al. 1998, McCann et al. 2003). La distinction entre effets extrinsèques (dus au climat ou à l'activité humaine) et intrinsèques (dus aux fluctuations naturelles du système, Lundberg et al. 2000), devient alors problématique car ils agissent souvent en synergie (Stromberg 1997, Rose 2004).

L'approche actuelle en écologie marine est donc intégrative et complexe, et privilégie les notions d'environnement, d'échelle, d'espace, de temps, en plus des interactions trophiques. L'avènement de nouveaux moyens d'observation du milieu marin (Sims 2003) est responsable pour une grande part de cette situation, de même que la fertilisation croisée entre différents champs de recherches (mathématiques appliquées, traitement du signal, océanographie physique...). Un rapide survol de la littérature nous montre ainsi que notre discipline est riche de méthodes étrangères, de concepts transverses, d'outils importés ou de techniques détournées de leur usage initial (Flierl et al. 1999, Dale et al. 2002)... La

complexification du problème écologique (l'interaction pêche-écosystème-climat) et sa difficulté avérée (Jonzén et al. 2002) nous laissent peu d'alternatives face aux signes de modification du climat et de la structure des écosystèmes toujours plus clairs et alarmants (Christensen et al. 1996, Petchey et al. 1999). Doit-on pour progresser (i) persévérer dans l'application des concepts actuels (paradigmes, modèles et méthodes d'analyse)?, (ii) espérer et fantasmer sur la mise au point de nouveaux outils ou réseaux d'observation?, ou (iii) proposer, encore, de nouveaux paradigmes ou de nouvelles approches au problème (Verity et al. 2002)?

Ces options sont examinées à l'échelle de cette thèse, mettant en scène une espèce remarquable parmi les scombridés, le thon rouge Atlantique (*Thunnus thynnus*). Le besoin de telles études se fait sentir par la pression élevée exercée par les pêcheries internationales sur son (ou ses) stock(s), les grandes incertitudes sur sa modélisation et sa gestion (ICCAT 2002) et plus généralement le manque de connaissances objectives sur sa biologie et son écologie (Fromentin 2003). On suppose en particulier un effet non négligeable du climat océanique et des conditions océanographiques locales sur sa distribution spatiale (Tiews 1978, Ravier and Fromentin 2004). Le sujet s'est donc construit à la croisée de l'écologie marine et de l'océanographie: le partage de mon temps entre la Direction Océanographie Spatiale de C.L.S. (Collecte Localisation Satellites) à Toulouse, et la Direction des Ressources Halieutiques à la station Ifremer de Sète a ainsi largement contribué à la diversité de ces travaux. Plusieurs questions ont été soulevées à leurs débuts:

- Quelle pouvait être l'effet de la variabilité océanographique sur la distribution du thon rouge?
- Existait-il un mode d'observation (en termes de résolution ou de système de mesures) privilégié pour étudier ces effets?
- Quels moyens d'analyse devait-on mettre en œuvre, voire mettre au point pour extraire l'information voulue de l'observation?
- Quelle forme pouvait prendre cette information?

La réflexion, d'abord totalement orientée sur l'analyse du signal océanographique et l'acquisition de connaissances sur le thon rouge, s'est peu à peu enrichie de questions plus fondamentales d'écologie quantitative.

Les deux axes principaux de ce travail sont: (i) la recherche et la construction de nouvelles approches pour l'étude de la dynamique spatio-temporelle du thon rouge, et de ses interactions avec son habitat, et (ii) une réflexion systématique sur la nature et les modalités d'extraction du signal écologique à partir de l'observation.

Quatre volets très divers constituent donc ce manuscrit: on débute, dans le chapitre 1, par un rappel des spécificités du milieu marin (par rapport à son homologue terrestre) et des difficultés inhérentes à son étude. La biologie et l'écologie du thon rouge sont rappelées, afin de circonscrire la problématique de ces travaux. Le chapitre 2 initie une réflexion sur les origines, l'étude et les effets sur le biotope des structures océanographiques comme les zones frontales. On y présente une étude sur la distribution spatiale **locale** de *Thunnus thynnus*, faisant intervenir des outils de télédétection (*i.e.* radiométrie satellite) et des observations de campagnes aériennes dans le Golfe du Lion. Ces travaux ont pour but d'apprécier l'effet des discontinuités océanographiques (*e.g.* fronts thermiques de surface) sur la répartition des bancs, de même que le besoin de méthodes spatiales adaptées au test d'hypothèse. Ces résultats sont accompagnés de deux manuscrits (dont un publié). La vision dynamique prônée par ces travaux a conduit à considérer un outil récent d'étude des animaux marins, la **marque archive**. Les aspects techniques et méthodologiques liés à cet outil sont rappelés au chapitre 3, où une technique de Monte-Carlo est proposée et testée pour extraire le signal des mesures archivées. La publication associée y est détaillée, ainsi que des méthodes particulières de traitement du signal. La **méthodologie** construite dans ces deux chapitres permet ensuite d'aborder sous un angle nouveau les fluctuations à long terme du thon rouge, telles que révélées par les captures historiques en Méditerranée. Cette analyse de **séries temporelles** est présentée au chapitre 4 sous forme d'une enquête, doublée d'un manuscrit, où méthodes classiques et exotiques se complètent pour extraire l'information écologique. On y mène en particulier une réflexion sur les difficultés et les possibilités offertes par le domaine fréquentiel en écologie quantitative. La conclusion prend la forme d'une synthèse de ces résultats, et du détail de quelques axes de recherche prometteurs en écologie marine. Les cinq articles et manuscrits complétant ce travail sont placés en annexe.

De l'enquête écologique

Notre discipline - l'écologie marine - se pose essentiellement en science de terrain, teintée de phénoménologie et basée sur le test d'hypothèse. Avec la découverte d'angles d'approches innovants, d'autres moyens de mesure et de nouvelles méthodologies d'analyse, notre représentation de la réalité écologique s'est complexifiée, amplifiant et ramifiant le problème initial (Wagner 1996). Plusieurs questions sur l'avenir de l'enquête écologique sont ici formulées.

Face à l'accumulation (inéluçtable) des savoirs écologiques et à la sophistication des outils d'observation et de prédiction des océans, il est tentant de vouloir adopter une approche intégrative, au nom des avancées de la systémique et de la théorie de la complexité (Von Bertalanffy 1968, Morin 1990). "L'idée océanographique" prend par exemple une place importante dans le processus de découverte en écologie marine. L'ajout d'un simple terme de forçage géophysique a eu de nombreuses implications pour notre vision du système biologique. Si ce concept n'est pas récent (Hjort 1914, Cushing and Dickson 1976, Steele 1989), il a été abordé de façons très diverses. Les méthodes d'écologie quantitative ont ainsi évolué de simples modèles 1D modulés par un forçage, à des modèles multi-dimensionnels complexes, autorisant la valorisation de données océanographiques de plus en plus riches (Lehodey et al. 1998, Siegel et al. 2003). Si l'objectif d'une telle approche est sans conteste primordial (*i.e.* traiter un système biologique dans sa globalité), le mode de résolution souvent choisi (le modèle comme pierre angulaire) est potentiellement limitant. En effet il appartient au scientifique d'établir une hypothèse nulle (H_0), donc un modèle de fonctionnement du phénomène caché, puis de l'évaluer, de le réfuter ou de l'accepter, pour ensuite l'améliorer (Popper 1982). Ce processus d'acquisition de connaissances est donc itératif. Or l'amélioration des moyens d'observations de la nature est constante: le dit modèle croît sans cesse en complexité. Le mode de progression classique (le test d'hypothèse itératif) devient donc de plus en plus difficile à appliquer, car le "modèle nul" (M_0) à construire devient à chaque fois moins parcimonieux et moins décomposable (de modèle nul M_0 , il devient "système nul" S_0 , un système étant par définition peu ou pas décomposable en sous-parties). Ainsi le degré de complexité du problème peut limiter *in fine* toute possibilité de test d'hypothèse (l'utilisateur devenant incapable de seulement formuler H_0). Puisqu'il ne véhicule pas la possibilité de réfuter ou de valider ses propositions, l'outil ainsi employé semble peu utile au raisonnement scientifique. Trop souvent, on vérifie un modèle en comparant ses sorties et les observations. Or on ne peut vérifier un modèle, mais seulement le valider. Ceci constitue une différence profonde. Lorsqu'un écart entre estimation et observation apparaît, il est aisément possible de corriger le modèle (par ajustement des paramètres) ou sa sortie (par recalage). En cela, les observations (donc la réalité terrain) deviennent subordonnées au modèle. Celui-ci est alors devenu extrêmement complexe: le coût de sa mise au point, on le comprend, peut freiner sa remise en cause s'il diverge. S'il reproduit avec succès certains aspects de la réalité, le problème est encore plus frustrant: quel paramètre, quelle équation en est alors responsable? Quel module doit-on conserver? Quelle sous-système doit être amélioré? Dans certains cas, il risque de perdre à la fois sa qualité de modèle (*i.e.* un outil de réduction de la complexité) et sa qualité de vecteur de connaissance scientifique (*i.e.* une déclaration toujours testable, donc réfutable). En d'autres termes, aborder un problème complexe de façon complexe peut constituer un investissement à risque.

Il est aujourd'hui reconnu que la compréhension des fluctuations des populations marines doit être pluridisciplinaire: est-ce en terme de types et de moyens d'observations, de formalisme ou de méthodologie? Est-il seulement possible de profiter de l'amélioration constante des techniques (d'observation ou de formalisation mathématique) tout en conservant la possibilité du test d'hypothèse?

1. Contexte océanographique et écologique

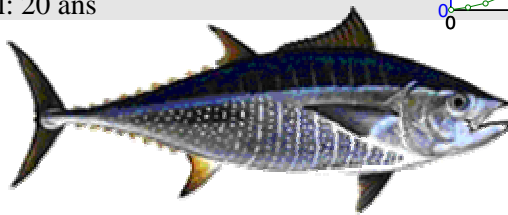
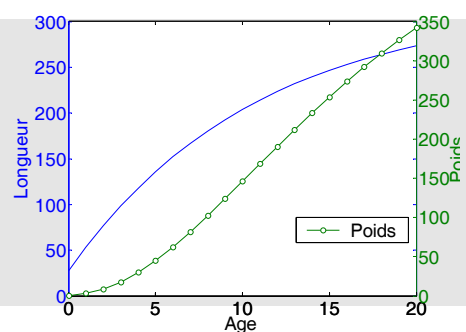
1.1. Le thon rouge Atlantique, *Thunnus thynnus*

Thunnus thynnus (Linné, 1758) est un Téléostéen, représentant de la famille des Scombridés. Il diffère du thon rouge du Pacifique (*Thunnus thynnus orientalis*) et du thon rouge du sud (*Thunnus maccoyii*) par quelques distinctions morphologiques mineures (Collette 1979). Quelques points clés de sa biologie et de son écologie sont résumés dans ces pages.

Carte de visite

(Farrugio 1981, Cort 1991, Mather et al. 1995, Fromentin and Fonteneau 2001)

- Espèce d'eaux tempérées
- Fenêtre de ponte: 1.5 mois/an
- Longueur à maturité: 115 cm
- Poids à maturité: 27.5 kg
- Age à maturité: 4.5 ans
- Taille maximale: 295 cm
- Poids maximal (record): 685 kg
- Age maximal: 20 ans



Biologie. Le thon rouge doit maintenir une vitesse de nage minimale pour assurer la ventilation branchiale et compenser sa flottabilité négative (Magnuson 1973, 1978). Il présente donc un métabolisme basal élevé, associé à un système cardio-vasculaire complexe autorisant un découplage thermique entre les organes "périphériques" (e.g. branchies) et les régions musculaires profondes (Korsmeyer and Dewar 2001). Les muscles natatoires opèrent ainsi à une température élevée (24 à 35°C), autorisant (i) leur relative indépendance vis-à-vis de la température ambiante, (ii) des incursions dans des masses d'eau froides pour une durée limitée, et (iii) des accélérations immédiates et intenses liés à la poursuite de proies ou à la fuite, jusqu'à 10 FL.s⁻¹ (Magnuson 1978, Wardle et al. 1989). Les viscères sont réchauffées de manière intermittente lors de l'ingestion et de la digestion de proies, augmentant la vitesse et le rendement de l'assimilation chez le thon rouge (Carey et al. 1984). On note que le cœur est placé en aval de l'échangeur thermique, ou *rete mirabilae*. Cette position "externe" oblige le muscle cardiaque à opérer à une température proche de celle du milieu, ce qui pourrait proscrire les changements thermiques brusques (lors de l'exploration verticale de la colonne d'eau, e.g. Brill 1984). Ceci n'est pas observé pour *Thunnus thynnus*, ou d'autres grands pélagiques comme *Thunnus obesus*

et *Xiphias gladius* (Carey and Robinson 1981, Lutcavage et al. 2000). Sa grande mobilité verticale vis-à-vis d'une thermocline marquée pourrait être liée à des augmentations temporaires du volume d'éjection systolique, compensant la bradychardie. Le maintien d'une forte oxygénation sanguine permet ainsi de pallier les besoins des muscles de type I (muscles rouges). Leurs propriétés aérobies autorisent une grande endurance et une nage soutenue (Lutcavage et al. 2000, Brill et al. 2002). Leur enveloppe est composée de muscles de type II (muscles blancs), de très fortes capacités aérobie et anaérobie: ceux-ci sont sollicités lors d'accélération intermittentes et fréquentes, et permettent une récupération rapide vis-à-vis d'une dette d'oxygène. Leur contribution à une stratégie de nage du type "propulsion-glissade" a été avancée (Weihs 1974, Videler and Weihs 1982). Un profil hydrodynamique caréné, une nageoire caudale d'aspect-ratio élevé associée à une grande raideur latérale complètent les capacités natatoires du thon rouge.

Le développement des larves (3-12 mm) est rapide: elles acquièrent leurs capacités natatoires dès 15 jours (Sella 1929). La croissance des juvéniles intervient essentiellement en été (50 à 60 cm l'année 1 puis 20 à 30 cm les années suivantes, dont 50% en été, Mather et al., 1995). La croissance se fait en longueur les trois premières années, puis concerne essentiellement la masse à partir de l'année 4 (Cort 1991). De nombreuses courbes de croissance ont été proposées dans la littérature (Farrugio 1981, Mather et al. 1995). Leur relative inconsistance a été attribuée (i) aux erreurs de mesure des otholithes et des coupes de vertèbres, ou (ii) à la disparité des origines géographiques des poissons capturés (i.e. favorabilité hétérogène des zones d'alimentation). La longévité du thon rouge est estimée à environ 20 ans.

Thunnus thynnus apparaît donc être remarquablement bien adapté à la vie pélagique, mais présente une morphologie et un métabolisme paradoxalement très gourmands en énergie. Sa grande capacité aérobie lui permettrait toutefois de limiter les conflits énergétiques (e.g. attribution aux différents postes: nage, récupération, assimilation, croissance, maturation des gonades).

Ecologie. La distribution spatiale du thon rouge couvre les eaux tempérées et tropicales de l'océan Atlantique, de l'Equateur aux îles Lofoten. Cette aire de répartition serait liée à la grande tolérance thermique des adultes (10-29°C, Carey and Gibson, 1983). Ceci est confirmé par les valeurs extrêmes mesurées par des marques archives (2-6°C, Block 2001, lutcavage 2001). L'Atlas des Pêcheries Thonières (Fonteneau 1997) confirme l'ubiquité de cette espèce, même dans les zones oligotrophes (e.g. le centre du gyre Nord Atlantique). Il évolue à des profondeurs de 0 à 500 mètres, avec des plongées enregistrées à 1000 mètres (Block and Stevens 2001). Les suivis par marques acoustiques ou

archives montrent qu'il passe l'essentiel du temps (90 %) dans les 50 premiers mètres (Lutcavage et al. 2000). Les profils de plongée ne semblent pas obéir à un déterminisme aussi clair que chez d'autres thons tropicaux. Le comportement grégaire varie avec l'âge: les bancs de juvéniles (50 à 100 individus) regroupent un nombre limité de classes d'âge. Les "géants" (2.50 à 3 m) se déplacent en solitaire ou en bancs de très petite taille (5-10 individus), groupés en tirailleurs, et sont observés en Atlantique Ouest (Lutcavage et al. 1996). Ces organisations grégaires seraient liées à la recherche de nourriture: pour assurer des rations journalières de l'ordre de 20 kg proie.jour⁻¹ (Crane 1936, 1936), des déplacements quotidiens de 70-80 km ont été observés. Les vitesses moyennes enregistrées par suivi acoustique sont de l'ordre de 1.5 à 5 km.h⁻¹, sans pouvoir différencier une stratégie de nage soutenue d'une progression de type "burst-and-glide" (Newlands et al. 2004). La prospection continue de l'habitat et l'alimentation semblent donc être essentiels chez le thon rouge: les contenus stomacaux le font apparaître comme un prédateur opportuniste (60% de crustacés, 20% de poissons et 20% de céphalopodes chez les juvéniles, Tiaws, 1962). Le spectre de taille des proies serait encore plus large chez l'adulte, et toucherait même les plus bas niveaux trophiques (gélatineux). Nous disposons de peu de données sur la fréquence d'ingestion de proies (Block and Stevens 2001). Ces mouvements trophiques à petite échelle (~1-100 km) doivent être différenciés des mouvements migratoires (~1000 km) liés à la recherche de conditions favorables à la croissance et au développement des œufs et larves. Les aires de ponte connues sont situées dans les mers méditerranées (Baléares, mer Thyrrénienne, golfe du Mexique) et apparaissent excessivement restreintes au regard de la distribution des adultes (cf. partie 1 de ce chapitre). Leur identification est actuellement associée à la présence soit i) d'œufs ou de larves de l'espèce, soit ii) de femelles matures prêtes à émettre les gamètes. Ces aires sont caractérisées par des phénomènes de convergence et de rétention liés à des tourbillons anticycloniques ou des zones frontales, potentiellement favorables à la survie et au développement des œufs et des larves (Bakun 1996, Garcia et al. 2003). Sa biologie reproductive est relativement mal connue. L'âge à maturité complète (100%) serait de 10 ans pour les individus en Atlantique ouest (~200 cm), et 5 ans en Atlantique est (~130 cm). L'hypothèse communément admise est celle d'une ponte annuelle de mai à juin en Méditerranée, avec une émission fractionnelle des gamètes dans des eaux de plus de 24°C (Schaefer 2001). De nombreuses aires de ponte secondaires ont été évoquées, dans la baie ibéro-marocaine, la zone canarienne, le golfe de Gascogne, le bassin Levantin en Méditerranée orientale (Karakulak et al. 2004) ou l'extension du Gulf Stream (Lutcavage et al. 1999). Leur fréquentation annuelle (i.e. retour au bercail, ou "homing") serait liée à la mémorisation d'une empreinte physico-chimique lors de la croissance larvaire (Cury et al. 1996). L'évaluation des zones de ponte secondaires ou hypothétiques reste à faire. Au vu des échelles de temps et d'espace concernées, le succès de telles migrations trans-océaniques semble dépendant d'un mécanisme de navigation hauturier. Si les déplacements locaux apparaissent comme non-aléatoires (Newlands et al. 2004) et obéissant à une intégration du

chemin parcouru (millinski 1994), l'orientation en migration pourrait être assurée par géomagnétisme (Carey et scharold, 1990, Walker, 1984, 1997) et olfaction à la côte (Atema, 1980). Une hypothèse encore peu développée est l'orientation par rapport au champ de lumière polarisée sous la surface, démontré comme hautement réminiscent de la polarisation atmosphérique (Horvath and Varju 1995, Deutschlander et al. 1999, Cronin and Shashar 2001, Phillips et al. 2001, Parkyn et al. 2003).

1.2. L'habitat océanique

Une courte description est ici fournie de quelques modes de variabilité de l'océan, de sa dynamique à long terme, *i.e.* de son climat. Les différences d'échelles spatiales et temporelles avec son homologue aérien (l'atmosphère) sont en particulier notées, et illustrées par l'exemple de l'**Atlantique Nord**. Les contraintes imposées aux espèces marines par ce milieu sont enfin évoquées.

Topographie. La partie nord de l'océan Atlantique est également divisée par une dorsale médiane, et est bordée par un grand nombre de mers adjacentes qui ont une grande influence sur la qualité de ses eaux (Tomczak et Godfrey 1994). Sa profondeur moyenne est de 3300 m, avec des plaines abyssales dépassant 5000 m. Des archipels (Açores, Canaries) et de nombreux monts sous-marins isolés ponctuent la partie Est du bassin. Le détroit de Gibraltar, de 22 km de large et de 300 m de fond, assure la communication avec la mer Méditerranée Eurafrique, dont la circulation est séparée en deux par le Déroit de Sicile. Le golfe du Mexique est connecté de façon multiple à l'océan, à l'est de Puerto Rico et entre Cuba et Haïti (profondeur : 1700 m), et entre la Floride et les Bahamas (profondeur : 750 m). Les mers adjacentes européennes (Mer du Nord et Mer Baltique) et la Mer du Labrador complètent ce tableau.

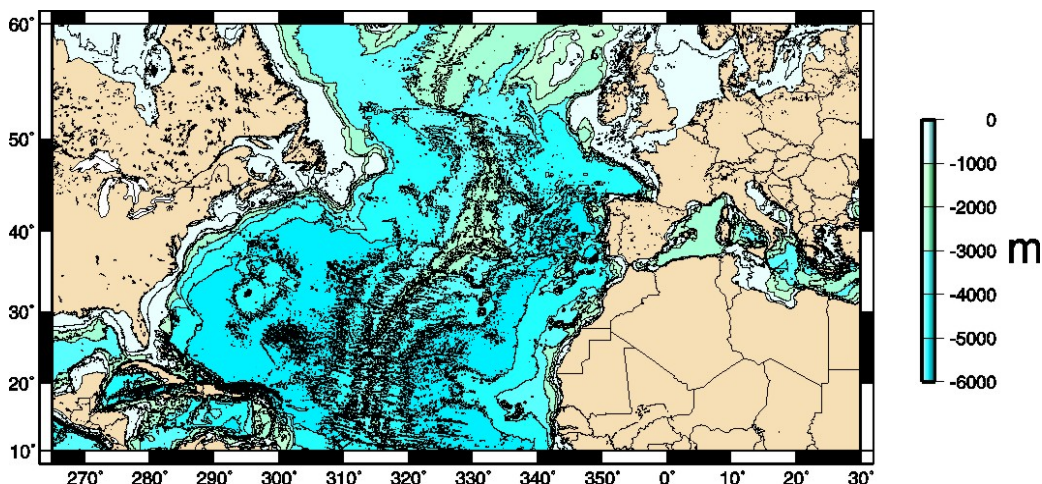


Figure 1-1. Bathymétrie de l'océan Atlantique Nord, dérivée de la base ETOPO2

Circulation. La circulation générale en Atlantique nord est constituée d'un gyre sub-tropical anticyclonique, bordé par le Gulf Stream et le courant du Labrador: sa dynamique est liée à la fois au forçage atmosphérique par les alizés, et à la circulation thermo-haline globale. L'activité tourbillonnaire, liée à la géostrophie, se fait sentir dans l'ensemble du bassin, en particulier dans la zone du Gulf Stream et de son extension. Au nord, le gyre polaire, cyclonique, est caractérisé par des eaux plus froides et plus denses: cette zone arctique est responsable de la formation de l'eau profonde atlantique.



Figure 1-2. Schéma de circulation thermo-haline en Atlantique Nord et mer Arctique (source: <http://oceanmag.whoi.edu/>)

En comparaison, la Méditerranée peut être vue comme un océan miniature, où opèrent les principaux mécanismes de la dynamique marine (e.g. forçage atmosphérique de la circulation superficielle, formation d'eau profonde en hiver, upwellings côtiers). Le climat sec et les faibles apports fluviaux en font un bassin de concentration, abritant plusieurs masses d'eau typiques (Millot 1999). La Méditerranée n'est réellement soumise au flux d'ouest atmosphérique qu'en hiver (avec des vents de nord violents vers le Golfe du Lion ou en Adriatique): le reste de l'année, la surface est forcée par des coups de vent locaux liés à la topographie. La circulation de surface produit des tourbillons de même échelle que leurs analogues océaniques mais généralement plus stationnaires ou de dynamique plus prévisible. Le flux sortant du détroit de Gibraltar génère un panache d'eau dense et salée, qui s'étend largement sur le fond océanique.

Echelles de variabilité. Les caractéristiques physiques du milieu marin le démarquent radicalement de son homologue terrestre: sa viscosité et sa capacité thermique intègrent et amortissent sensiblement la variabilité des forçages atmosphériques (Steele 1985). Les temps de renouvellement liés à la circulation intra- et inter-bassins génèrent des fluctuations à grande longueur d'onde ($\sim 10^1$ - 10^3 ans, $\sim 10^3$ - 10^5 km). Il est aussi admis que la conformation des bassins et l'existence de différents modes de variabilité (e.g. l'instabilité barocline, ondes de marée interne)

interagissent pour générer de nombreux signaux de période stable. Le spectre de variabilité océanique est donc complexe, et peut varier de façon régionale: les mers adjacentes, de par leurs dimensions, auront ainsi leurs propres spectres de réponse aux forçages atmosphériques, eux-mêmes dépendants des conditions climatiques globales. L'Oscillation Nord-Atlantique (ou NAO) est ainsi responsable d'une grande part des fluctuations atmosphériques et océaniques côtières en Atlantique Nord-Est et en Méditerranée (Hurrell et al. 2001)

A toutes les échelles, de nombreux autres mécanismes altèrent directement ses propriétés (e.g. réchauffement de surface, évaporation, précipitation, production primaire, échanges gazeux océan-atmosphère). L'advection et le mélange sont généralement prédominants, au détriment de la diffusion moléculaire. Une multitude de traceurs biogéochimiques sont soumis à ces processus (e.g. température, salinité, oxygène et dioxyde de carbone dissous, matières phosphatées, nitrates ou carbonées...): si leur distribution est naturellement corrélée à l'advection aux échelles $\sim 1-10^2$ km (Gower et al. 1980), leurs différentes échelles de temps caractéristiques contribuent à fortement structurer le milieu océanique (Gower et al. 1980, Steele 1989, Mahadevan and Campbell 2002).

Cette structuration se révèle inégale dans les trois dimensions: la finesse de la couche d'eau couvrant le plancher océanique, la rotation de la terre et la stratification en densité font que les flux horizontaux dominent les flux verticaux, malgré des gradients plus prononcés dans cette dimension. Ces flux verticaux sont toutefois essentiels à la productivité primaire des océans, car les remontées de nutriments dans la couche euphotique en dépendent. Cette diversité de mécanismes et de patterns impliquent que la variabilité océanique est mieux appréhendée selon un continuum d'échelles, et se plie mal à la classification¹ (Steele 1985, Longhurst 1998).

Impacts écologiques. D'un point de vue biologique, on s'attend à ce que les constantes de temps de la physique (e.g. phénomènes advectifs, turbulence) et de la productivité primaire soient largement corrélés (Levin and Segel 1976, Legendre and Demers 1984, Steele 1995). Par rapport au terrestre, cette production primaire marine est donc relativement moins prévisible (*i.e.* moins stable) dans l'espace et le temps, posant un défi à la fois aux organismes marins et à l'écologiste tentant d'inférer un mécanisme à partir d'observations. Pour illustration, le temps de résidence du carbone organique est 140 fois plus élevé sur les continents que dans l'océan (Cohen 1994).

¹ "The ocean is a turbulent fluid in intimate contact with another turbulent fluid, the atmosphere. Although I am unaware of any formal theorems on the subject, experience with turbulent systems suggest that it is very unlikely that any components of such a complex nonlinear system can actually remain fully steady; ... the frequency/wavenumber spectrum of the ocean circulation is almost surely everywhere filled" (Wunsch 1992)

L'organisation spatiale de la productivité primaire océanique est assez bien comprise: en bordure Est, des upwellings côtiers ou liés aux archipels (e.g. au Portugal, Maroc et Canaries) offrent la principale source de production primaire dans la zone épipélagique. Le gyre polaire, plus froid et largement sujet au forçage éolien est lui aussi très productif: la faible stabilité des couches de surface y limite la production hivernale, et donc le contrôle par le compartiment zooplanctonique lors la floraison printanière (Evans and Parslow 1985). Le centre du gyre sub-tropical est lui relativement oligotrophe, du fait d'une plongée générale de la pycnocline. Plus localement, les remontées d'eau liées à la dynamique géostrophique génèrent des enrichissements locaux, en général moins prévisibles. L'organisation verticale de cette production reste mal connue, de même que ses interactions avec des phénomènes intermittents comme les fronts de surface.

A l'échelle des communautés marines, ce biotope est donc tout sauf statique et monotone. L'habitat océan impose de nombreuses autres contraintes: c'est d'abord un milieu visqueux, où le mouvement est d'autant plus coûteux que l'animal est de taille réduite. Les faibles nombres de Reynolds prévalent aux échelles cellulaires, et font de la turbulence un phénomène critique pour la nutrition des petits organismes (Pedley 1997). C'est un milieu dense, donc porteur, qui facilite le déplacement par advection de petits organismes du nekton, mais aussi la migration trans-océanique des plus grands poissons et mammifères, car le déplacement d'un surplus de masse représente un surcoût énergétique faible. Le mouvement dans le milieu marin est donc un processus fondamentalement différent du mouvement sur les substrats terrestres ou dans l'atmosphère.

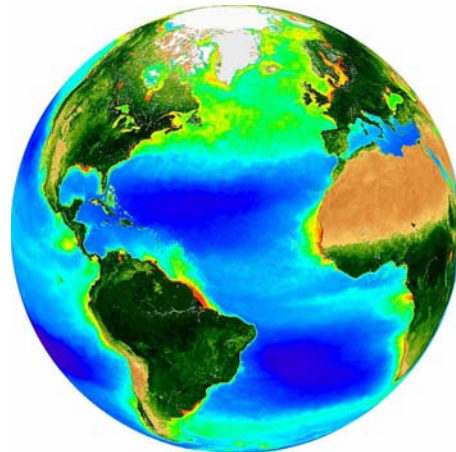


Figure 1-3. Distribution climatologique de pigments chlorophylle-a océanique et Index de Végétation terrestre (printemps boréal) "Design your own SeaWiFS globe" - http://oceancolor.gsfc.nasa.gov/cgi/biosphere_globes.pl

La majorité des espèces marines passe aussi par un stade planctonique aux premiers âges (*i.e.* œufs puis larves) et sont directement soumises à la variabilité du milieu: ceci est, encore, sans équivalent en terrestre. Ce découplage (en terme de vulnérabilité aux perturbations) des phases larvaires et adultes limite l'applicabilité de concepts traditionnels

type "r-K" (Steele 1985, 1995). Même aux stades juvéniles ou adultes, le concept d'homéostasie est relatif, tant chimiquement (e.g. l'excrétion des sels) que thermiquement (e.g. la masse d'eau formant un puits de chaleur de capacité quasi infinie).

Enfin, les contraintes hydrodynamiques et la pression de prédation semblent avoir conduit à une variété restreinte de morphotypes pélagiques (*i.e.* corps musculeux, fusiformes) mais à une infinité d'adaptations comportementales (e.g. bancs, migrations). Ces processus, intervenant à l'échelle de l'individu, sont propagés par cascade jusqu'au niveau des populations, voire interagissent avec ce dernier niveau: la délimitation d'entités, d'écotones, de métapopulations est alors moins aisée qu'en milieu terrestre, plus structuré, plus statique, plus propice à l'observation. En d'autres termes, il apparaît plus difficile de relier les patterns observés en milieu marin à des processus similaires à ceux postulés pour le milieu terrestre (Steele 1985, Verity et al. 2002). Les interactions des processus biotiques avec les propriétés du fluide offrent de multiples options aux populations marines, et tendent en même temps à lisser et faire converger des patterns d'origines diverses (c'est un problème d'équifinalité). L'approche phénoménologique et ses dérivées (e.g. la mise en corrélation) sont donc fortement indiquées² mais doivent être employées avec précaution.

En résumé, il existe de nombreuses origines potentielles aux fluctuations des populations marines, de la circulation thermo-haline à grande échelle (e.g. les anomalies de salinité) aux phénomènes locaux (e.g. les systèmes frontaux), en passant par le climat régional (e.g. les effets de la NAO sur les communautés de mer du Nord)(Hurrell 1995). Leur quête a permis d'expliquer de nombreux cas de variations d'abondance, de translations spatiales ou de changements trophiques dans les écosystèmes exploités (Aebischer et al. 1990, Fromentin and Planque 1996, Heyen et al. 1998, Planque and Taylor 1998, Ottersen et al. 2001). Le spectre de variabilité océanique fournit donc aux compartiments biotiques un grand nombre d'options, et structure leurs fluctuations spatiales et temporelles de façon remarquablement similaire (Haury et al. 1978). Si les oscillations de très grande longueur d'onde ont peu de chances d'être intégrées par les populations marines (*i.e.* elles seront seulement propagées, linéairement ou non), des réponses adaptées, ou non triviales, peuvent être attendues pour des fréquences spatiales ou temporelles plus courtes, compatibles avec les processus écologiques (Steele 1985). Cette vision, dans le cadre d'une approche évolutionniste, permet d'imaginer une cohérence entre les traits d'histoire de vie d'une espèce (et ses adaptations comportementales) et la variabilité biogéophysique de son habitat (e.g. les systèmes océaniques où la régénération des nutriments semble favoriser une dominance des flagellés et gélatineux, contre les chaînes trophiques dominées par les

² Le risque est de détecter des relations statistiquement significatives mais écologiquement improbables ou difficiles à vérifier, alors que l'approche mécaniste (construction d'un modèle basé sur la théorie appropriée) proposera des solutions écologiquement adéquates mais difficiles à tester (Jonzén et al. 2002).

diatomées, plus récentes et classiquement employées dans la vision "bottom-up") (Parsons 1979, Rigby and Mislom 2000).

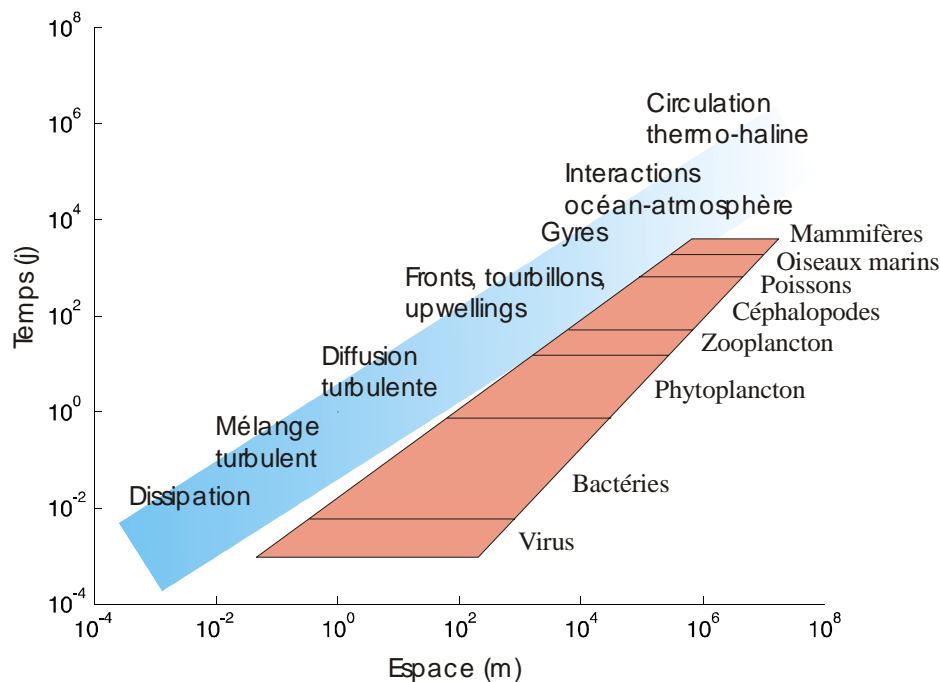


Figure 1-4. Echelles spatio-temporelles dominantes pour la mécanique océanique et les communautés biologiques. D'après Hoffman et Powell (1998).

Dans cet habitat hétérogène et dynamique, apparemment oligotrophe, les grands prédateurs pélagiques comme les thonidés font face à un défi d'ordre énergétique et reproductif: la survie de l'individu est liée à la satisfaction de très hauts besoins énergétiques (Block and Stevens 2001), alors que la survie de l'espèce dépend e.g. de l'adéquation entre les fenêtres de ponte et les conditions ambiantes (Cushing and Dickson 1976, Fromentin and Fonteneau 2001). Les dimensions spatiales et comportementales prennent ici une importance majeure pour ces espèces hautement mobiles. La perception de l'habitat et les modalités de son exploration deviennent ainsi des questions importantes, appelant à de nouveaux paradigmes (Sims 2003). Exploités pour leur valeur économique, les thonidés sont le sujet de connaissances inégales selon l'espèce: si la physiologie, l'écologie et l'éco-éthologie des espèces tropicales sont relativement bien connues, celles des thons tempérés, et plus particulièrement le thon rouge comportent de larges zones d'ombre.

2. Fronts océaniques: une écologie des interfaces

L'océan est un système globalement stratifié montrant une forte asymétrie entre ses dimensions horizontales et verticales. Les processus biologiques interviennent essentiellement dans sa couche supérieure euphotique (Lévy et al. 2001), dont l'épaisseur varie de façon saisonnière et dans l'espace: les discontinuités verticales (e.g. pycnoclines, thermoclines, haloclines...) et les structures tourbillonnaires font l'objet d'une riche littérature et ne seront pas abordées ici (The Ring Group 1981, Robinson 1983, Kamenkovich et al. 1987).

Nous nous restreindrons aux discontinuités horizontales comme les **fronts**, relativement moins bien connus en océanographie biologique. Leur effet sur la structuration et la productivité de l'habitat pélagique reste en particulier mal identifié (Martin et al. 2002, Etnoyer et al. 2004). Après le rappel des **mécanismes** intervenant dans leur formation, les méthodes de **détection** et de **cartographie** des fronts sont présentées. L'impact potentiel des fronts sur le **biotope** est ensuite abordé, en précisant leur influence sur la distribution spatiale de **prédateurs** comme *Thunnus thynnus*. La **méthodologie** de la mise en relation des mesures de physique/biologie est discutée, ainsi que les problèmes induits par leur structure spatiale. Les manuscrits 1 et 2 viennent illustrer ce tour d'horizon des structures frontales de surface.

2.1. Eléments d'océanographie frontale

Définitions. 1) Le terme de front océanique est habituellement employé pour décrire une zone où le **gradient** d'une variable géophysique, e.g. température, salinité ou densité, est notablement plus élevé qu'une certaine valeur caractéristique. Cette valeur peut être par exemple le gradient climatologique dans la zone étudiée (Fedorov 1986). Souvent, un champ de température est considéré présenter des fronts pour des valeurs $\Delta SST \geq 0.5-1.0^\circ C/km$. 2) On peut aussi définir un front comme une **interface** séparant deux masses d'eau homogènes mais de caractéristiques différentes (Owen, 1981).

Mécanismes. Les principaux mécanismes de la « frontogenèse » sont: (i) un processus hétérogène de **mélange** interrompant la stratification de la colonne d'eau (en général dû à une augmentation de la vitesse et donc de la turbulence et de ses effets), (ii) une **instabilité** du profil de densité (dû e.g. à la topographie, un refroidissement ou une évaporation superficielle), (iii) une **confrontation** de masses d'eau d'origines géochimiques différentes (e.g. flots issus de bassins ou de mers adjacentes, remontées d'eau froide par upwelling). La densité de l'eau de mer est une fonction de sa température et de sa salinité: un gradient maximal en densité génère un front dit **front de densité**, associé à un fort courant (ou jet) longeant les isopycnes, et à un flux convergent (Pollard and Regier 1992). Un tel front possède une forte

énergie potentielle et donc une persistance notable. Un front où le gradient de température compense la variation de salinité est dit **compensé en densité**: son énergie potentielle est moindre et il est plus sensible aux forçages locaux et régionaux. Il est donc souvent caractérisé par de nombreux méandres et interdigitations.

Classification élémentaire. (i) Les **fronts planétaires** structurent l'océan à l'échelle de bassins entiers (e.g. les Fronts Subtropicaux formés à la convergence de la couche d'Ekman, les Courants Circumpolaires et les interfaces entre les Gyres Subtropicaux et Subpolaires). Ils sont quasi-permanents et leur impact sur les échanges océan/atmosphère est non-négligeable (Tomczak et Godfrey, 1994). (ii) Les **fronts d'upwellings** se forment lorsqu'une thermocline émerge en surface sous la pression d'un flux vertical: ils sont associés en côtier au mouvement vers le large de la couche d'Ekman, et suivent la pulsation saisonnière du forçage par les vents (Johns et al. 1992). En mer ouverte ils sont essentiellement associés aux remontées d'eau dues à la géostrophie. (iii) Les **fronts de talus** sont formés par le gradient de pression entre les masses d'eau océanique et l'eau occupant le plateau continental, qui entretient ainsi un courant parallèle aux isobathes. Ces fronts sont donc plus stationnaires que les autres types élémentaires, et leur localisation varie seulement lors de la formation ponctuelle de méandres et de tourbillons. Le rayon de déformation de Rossby donne l'étendue de ce type de fronts (Tomczak et Godfrey, 1994). (iv) Les **fronts de mer peu profonde** sont parmi les plus connus, et correspondent à l'augmentation critique de la vitesse des courants de marée avec la remontée du fond en côtier: si la turbulence est suffisante pour détruire la stratification de la colonne d'eau, un front se forme entre la zone stratifiée et l'eau mélangée à la côte. L'énergie nécessaire pour cela est proportionnelle à h/u^3 , où h est la hauteur d'eau et u la vitesse du courant de marée: les fronts côtiers sont donc prédictibles dans une large mesure (Simpson and Hunter 1974). (v) Les **fronts de cisaillement** apparaissent à l'interface de masses d'eau de vitesses différentes, ou lorsque les vecteurs courants sont parallèles au gradient spatial en température ou salinité. Ces fronts présentent des phénomènes de turbulence et des mouvements verticaux intenses (Prieur 1985, Belkin and Cornillon 2003). (vi) Les **fronts d'estuaires** sont associés aux panaches fluviaux (l'eau douce formant une couche moins dense en surface, Largier 1993) et sont largement soumis aux forçages atmosphériques et aux intrusions de courants sur le plateau continental (Millot 1990).

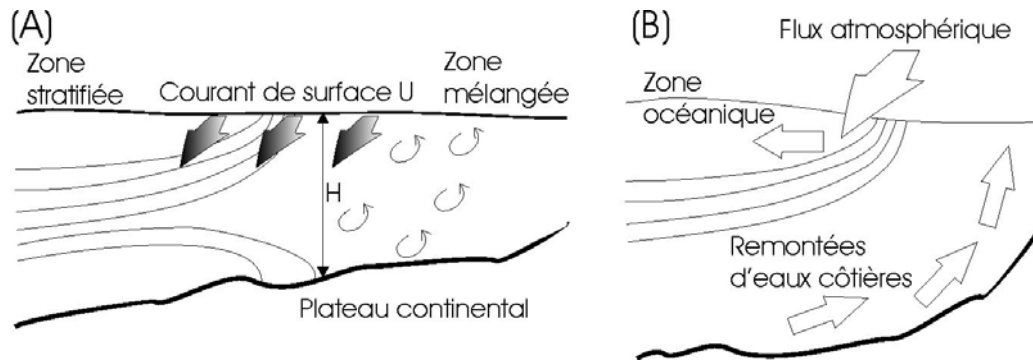


Figure 2-1. Exemples de mécanismes de la frontogénèse. A) fronts de marée. B) fronts d'upwelling

Interaction avec les traceurs océaniques.

L'advection par les courants est une source importante d'hétérogénéité (Gower et al. 1980). Chaque traceur possède un temps caractéristique (e.g. d'équilibre entre quantités dissoutes et gazeuses, ou de doublement de population pour une entité biologique) qui régit la variabilité spatiale de son champ de concentration. La diversité de ces échelles caractéristiques et des sources potentielles contribue à faire de la surface océanique un milieu structuré, ou "patchy". Par extension, les zones de fort gradient peuvent être là encore considérées comme des fronts. L'interaction avec les fronts "classiques" thermo-halins doit aussi être considérée: leur persistance dans le temps et l'espace conditionnera la dynamique de variables comme la matière organique dissoute, le dioxygène dissous...

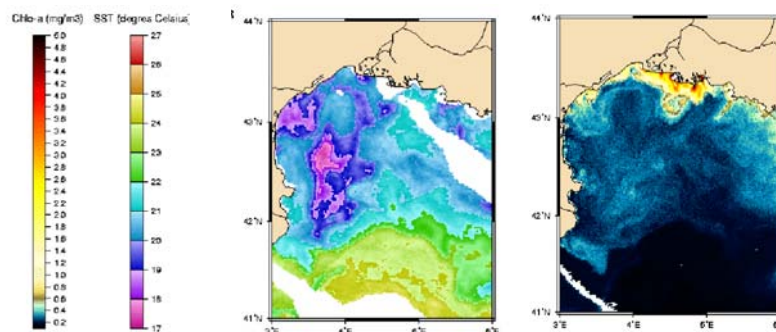


Figure 2-2. Images de température de surface AVHRR 2km et SeaWiFS 1.1 km dans le Golfe du Lion, au 21 septembre 2000

L'hétérogénéité spatiale se forme donc au croisement de processus tendant à lisser leur distribution (e.g. par mélange, diffusion, équilibre avec la phase gazeuse) et ceux renforçant les gradients (e.g. upwellings intermittents). On a ainsi montré, dans ce dernier cas, qu'un schéma de remontée d'eau spatialement hétérogène favorisait la redistribution horizontale de nutriments dans la zone euphotique, et donc la production globale par rapport à un schéma homogène (Martin et al., 2002). Quantifier la variabilité spatiale des différents traceurs océaniques est devenu d'une grande importance, comme en témoigne la riche littérature

sur le sujet (Longhurst 1995, Lovejoy et al. 2001, Mahadevan and Campbell 2002, 2003, Doney et al. 2004)

Observation et détection. Les zones de gradient ont fait l'objet de très nombreux rapports en tant que structures remarquables lors de campagnes en mer (Fedorov 1986). La faible couverture spatiale de ces mesures a longtemps restreint l'observation aux fronts de courants de bordure et autres structures quasi-pérennes. Depuis 1985, l'océanographie spatiale fournit une cartographie globale de la température de surface grâce au radiomètre AVHRR. CZCS (1989-1992) et SeaWiFS (1997-?) fournissent une estimation de la chlorophylle-a intégrée sur la couche euphotique. Ces deux traceurs (SST et Chlo-a) sont connus pour être largement corrélés, car sujets aux mêmes phénomènes advectifs (Denmann 1976, Hood et al. 1990, Barnard et al. 1997). La technologie SAR fournit aussi une information sur les fronts, spécialement en côtier où l'opposition stratification/mélange possède une forte signature, la turbulence de surface favorisant la diffraction du signal. Pour ces applications, l'**analyse d'images** fournit des outils adaptés à la détection des fronts. C'est sans étonnement que l'on constate l'emploi intensif d'opérateurs "gradient" pour la détection des fronts (Kazmin and Rienecker 1996, Yuan and Talley 1996, Ullman and Cornillon 2000, Wang et al. 2001, Nakamura and Kazmin 2003, Etnoyer et al. 2004). Le calcul consiste en une simple différence finie entre pixels voisins en X et en Y ou, pour les puristes, une convolution de l'image par les opérateurs de Sobel ou de Prewitt. Conjuguée à un lissage préalable et à un seuillage par hystérésis pour connecter les pixels orphelins, cette étape constitue la base de l'algorithme de Canny, un grand classique de ce domaine (Canny 1986).

Cet algorithme a été implémenté et appliqué dans le cadre de l'article 2: si les performances étaient satisfaisantes pour les besoins de cette étude, il apparaît que cette méthode simple manque de robustesse en cas de bruit instrumental ou atmosphérique, et aux abords des masques de nuage. Elle manque aussi de pertinence océanographique (un front étant aussi et surtout une **interface** séparant deux masses d'eau). Un algorithme traitant explicitement les interfaces d'une image a été proposé par Cayula et Cornillon (1992). Il consiste en une fenêtre glissante dans laquelle la bimodalité de l'**histogramme** est testée, puis une valeur optimale de coupure est estimée. La région est ensuite séparée en deux populations de pixels: la **cohérence** de chacune de ces masses d'eau est alors vérifiée. Enfin, chaque pixel marqué comme front est soumis à une analyse du gradient classique. Cette version de l'algorithme a été ici implémentée sous Matlab et employée à des fins de cartographie de fronts thermiques en Atlantique Nord-Est et Méditerranée (article 1). Cet algorithme a bénéficié dans la littérature d'une suite d'améliorations: il a été étendu à l'analyse séquentielle d'images (Cayula and Cornillon 1995), la taille de la fenêtre glissante peut être rendue variable et coïncidante avec la portée du variogramme (Diehl and Wells Budd 2000), ou il peut opérer sur plusieurs canaux de radiométrie (Miller 2000). Des raffinements supplémentaires peuvent être ajoutés pour mimer

au plus près le travail de l'expert sur l'image: une rotation de la fenêtre d'échantillonnage pour échantillonner au mieux un front oblique, une méthode de Monte-Carlo préservant l'autocorrélation du champ de température pour le test du gradient (Fortin and Jacquez 2000), ou encore l'ajout du seuillage par hysteresis pour connecter les pixels. La totalité de ces étapes est synthétisée en Fig. 2-3, et constituerait un algorithme statistique de détection des bordures (Fagan et al. 2003), pour des utilisations tant terrestres que marines. En d'autres termes, un tel algorithme ne reposerait plus sur un *a priori* fourni par l'utilisateur, mais assurerait la détection statistique objective et adaptative des discontinuités.

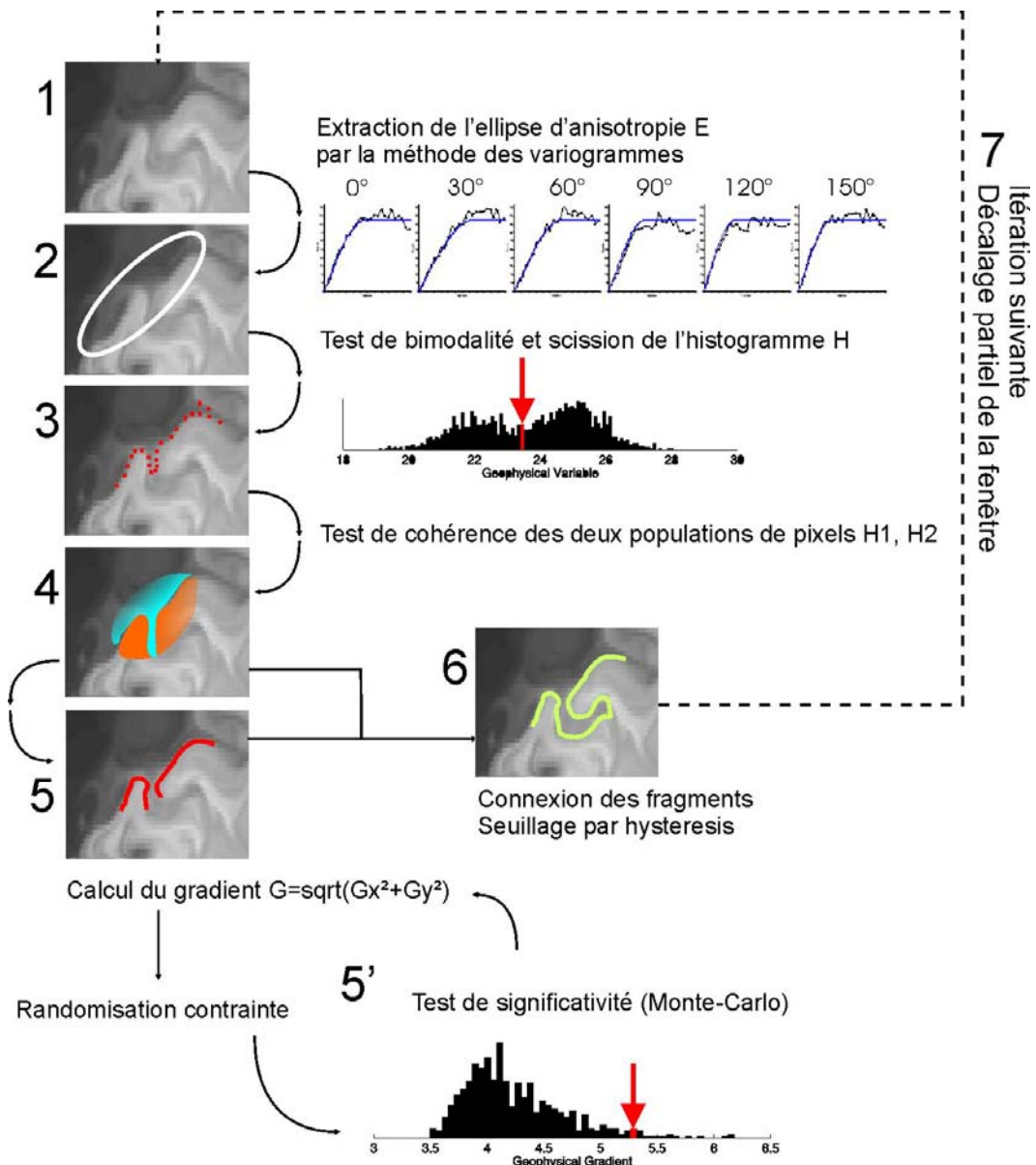


Figure 2-3. Proposition d'un algorithme de détection des fronts intégralement statistique, basé sur le calcul de l'histogramme (Cayula et Cornillon, 1992).

Cartographie. De tels algorithmes s'avèrent gourmands en temps de calcul: l'amélioration des moyens informatiques et la taille croissante des bases de données a permis l'initiation de cartographies frontales régionales et globales (Ullman and Cornillon 1999, Belkin and Cornillon 2003, Belkin et al. 2004). Ces travaux portent sur la quantification objective de la stabilité spatiale et de la persistance des fronts océaniques, qu'ils soient associés à des courants de bordure, des upwellings côtiers, ou des phénomènes intermittents liés à la géostrophie.

Pour enrichir notre propos (*i.e.* l'étude de leur rôle dans la structuration de l'habitat pélagique), une cartographie des fronts thermiques de surface a été conduite pour l'Atlantique Nord-Est et la Méditerranée (manuscrit 1). Une base de données à haute résolution de fronts a été constituée à partir d'images SST LAC (2 km/pixel) traitées par le Centre de Météorologie Spatiale de Lannion (Anonyme 2002). Chaque image a ensuite été analysée en employant l'algorithme décrit par Cayula et Cornillon (1995). Le signal (*i.e.* la densité de front comme une probabilité de présence) a ensuite été décomposé en fonctions empiriques orthogonales (EOF) spatiales et temporelles, suivant la méthode de Lanczos (Toumazou and Cretaux 2001). Cette méthode s'apparente à une Analyse en Composantes Principales et révèle comment des patterns récurrents s'additionnent et se complètent pour générer le signal environnemental (on notera que la validité physique des patterns mis en évidence par EOF n'est pas toujours assurée). La cartographie trimestrielle de quatre zones (Portugal/Canaries, Gascogne/Irlande, Méditerranée occidentale et orientale) illustre quatre des cinq grands types de fronts précédemment évoqués: (i) les fronts liés aux upwellings (côtes portugaises et marocaines), (ii) les fronts liés aux talus continentaux (nord Gascogne, nord-ouest Irlande), (iii) les fronts liés à la marée (Manche, mer d'Iroise et nord Irlande), (iv) les fronts liés aux panaches fluviaux (Rhône). Les effets d'îles (Canaries, Açores) ont une signature particulièrement visible dans cette cartographie. La décomposition en EOF a permis la décomposition du signal, l'extraction de la saisonnalité et son attribution à des régions bien définies: les quatre régions partagent le même pattern saisonnier, avec un premier axe important (~30-50% de variance) d'amplitude maximale en juin-août. L'inspection du premier mode spatial permet d'associer ce phénomène à l'action duelle des forçages atmosphériques: la pulsation des upwellings de bordure Est-Atlantique est maximale en été, alors que les flux d'ouest affectent l'ensemble des zones en hiver, régularisant les structures de surface. La décomposition permet aussi d'apprécier les facteurs prédominants de la frontogenèse: les côtes atlantiques françaises sont sous l'influence principale de la marée, alors que le signal dominant au Portugal et au Maroc est lié aux fronts délimitant les upwellings. La méditerranée occidentale présente une structuration marquée liée à sa circulation cyclonique (*e.g.* courant Nord-Algérien, courant Liguro-Provençal), et plus localement à l'orographie (*e.g.* bouches de Bonifacio). Le bassin oriental apparaît beaucoup plus bruité, avec des événements frontaux apparemment liés aux vents de Nord.

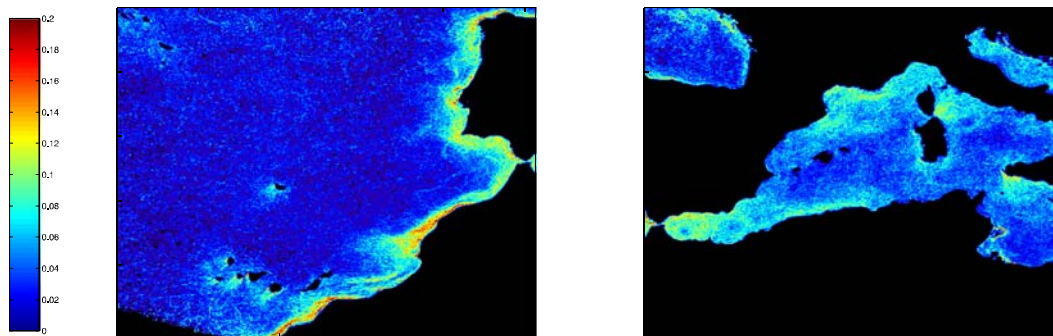


Figure 2-4. Exemples: densités moyennes de fronts thermiques sur le trimestre juillet-août-septembre dans la zone canarienne (à gauche) et la Méditerranée occidentale (à droite), calculées sur 2001-2004 à partir des images SST 2km du CMS de Lannion.

Les EOFs – Empirical Orthogonal Functions

Soit \mathbf{X} une matrice de mesures de dimension (m,n) (m étant une dimension spatiale et n une dimension temporelle). La Décomposition en Valeurs Singulières (SVD, *e.g.* Lawson et Hanson, 1974) permet d'écrire \mathbf{X} sous la forme:

$\mathbf{X}=\mathbf{U}\mathbf{S}\mathbf{V}^T$, où \mathbf{U} contient les EOFs spatiales de dimension (m,r) , \mathbf{V} les EOFs temporelles de dimensions (n,r) et \mathbf{S} contient les valeurs singulières ρ de dimension (r,r) (r étant le rang $r\leq\min(m,n)$). En pratique, \mathbf{X} est centré pour travailler sur les anomalies du signal.

Les EOFs sont donc analogues à une Analyse en Composantes Principales d'une matrice temps-espace (Legendre et Legendre 1998). Cette méthode de décomposition permet de conduire des analyses de variabilité (*e.g.* Hendricks et al. 1996), de filtrer, de comparer modèles et données (Brachet, comm. pers.), ou réduire la complexité dans les étapes d'assimilation...

Modélisation statistique. La cartographie des fronts fournit de précieuses informations sur la variabilité de surface, mais elle reste toutefois insatisfaisante: elle suppose en effet un degré minimal de leur **persistance**. De fait, les fronts semi-stationnaires (*e.g.* fronts de talus, de marée...) seront les mieux décrits dans leur extension spatiale et temporelle par des cartes de densité. Les fronts **transitoires** et mobiles liés à la circulation géostrophique ou à des upwellings locaux/intermittents sont donc lissés et **sous-représentés** dans l'agrégation d'images successives. Quelle serait donc la prochaine étape à aborder dans la description objective des fronts? Une **cartographie dynamique** pourrait par exemple reposer sur l'ajustement d'une fonction de covariance explicitant les phénomènes de propagation (Schaeffer et al. 2002):

$$C(r,t)=f(r).\exp\left(-\frac{t^2}{T^2}\right), \text{ où } r=\sqrt{\left(\frac{dx-C_{px}dt}{L_x}\right)^2+\left(\frac{dy-C_{py}dt}{L_y}\right)^2}$$

où f est une fonction de type exponentiel, et T , L_x et L_y sont les échelles de temps et d'espace. C_{px} , C_{py} sont les vitesses de propagations zonale et méridienne. La séparation des modes stationnaires et propagés s'effectue ensuite en estimant et cartographiant ces termes C_{px} , C_{py} . Intuitivement, la résolution spatiale et temporelle doit être suffisante pour résoudre la propagation des fronts, sans être limitante en termes de densité de données (les effets de la nébulosité devenant plus sensibles avec la

précision de la résolution). La reconstruction des champs de température doit alors être envisagée (e.g. par analyse objective), en privilégiant les méthodes préservant les gradients de la variable étudiée (i.e. un phénomène typiquement non-linéaire). Une méthode récente de reconstruction non-paramétrique (Beckers et Vixen, 2003), basées sur les EOFs, apparaît être un bon candidat.

Les EOFs comme alternative à l'analyse objective et au krigage

La cartographie avec données manquantes repose souvent sur le calcul d'une fonction de covariance empirique $C(x_\alpha-x_\beta)=C_{\alpha\beta}$ et sa modélisation (Matheron 1971, Bretherton et al. 1976). La résolution d'un système de **krigeage** permet alors l'estimation de la variable en tous points X_0 non échantillonné:

$$X^*_0=m+\sum_{\alpha=1}^n\lambda_\alpha(X_\alpha-m)$$

où m est la moyenne spatiale, X_α sont des mesures du voisinage, λ_α sont des poids tels que X^*_0 est sans biais, et l'erreur $X^*_0-X_0$ est de variance minimum. Ce système se résout en cherchant à minimiser l'erreur de prédiction:

$$\text{Var}(X^*_0-X_0)=E\left[\sum_{\alpha=1}^n\lambda_\alpha(X_\alpha-m)-(X_0-m)\right]^2=\sum_{\alpha=1}^n\sum_{\beta=1}^n\lambda_\alpha\lambda_\beta C_{\alpha\beta}+\sigma_0^2-2\sum_{\alpha=1}^n\lambda_\alpha C_{\alpha 0}$$

La dérivée d'ordre 1 s'annule pour $\sum_{\beta=1}^n\lambda_\beta C_{\beta\alpha}=C_{\alpha 0}$

En pratique, cette méthode s'avère peu efficace pour des jeux de données de grande dimension (Schaeffer, comm. pers.): un système réduit peut être obtenu en ne travaillant que sur les axes de variabilité principaux de la variable (e.g. par EOF), tout en conservant la structure de la corrélation spatiale et temporelle (Holbrook and Bindoff 2000). De plus l'estimateur ainsi obtenu n'est optimal que pour des dynamiques linéaires.

L'application directe des EOFs à une matrice temps-espace incomplète permet d'aborder le problème sous un angle nouveau: en retirant la moyenne de l'échantillon m^* (assimilée dans la suite à la moyenne vraie) puis en égalant chaque valeur manquante X_{ij} à zéro, on obtient un jeu de données *a priori* non biaisé (si $E[m^*-m]=0$). Une décomposition en EOF en n axes fournit alors une estimation de ces valeur manquantes (grâce aux k premiers axes):

$$X_{ij}=\sum_{p=1}^k\rho_p(u_p)_i(v_p^T)_j, \text{ où } k < n$$

Cette estimation est ensuite réinjectée dans une décomposition en EOF, jusqu'à obtenir la convergence des valeurs initialement manquantes. L'évaluation du nombre optimal k d'axes nécessaires à la reconstruction est décrit par (Beckers and Rixen 2003). Cette méthode offre l'avantage de ne pas reposer sur une coupure du spectre fréquentiel de la donnée ou, de façon équivalente, un lissage de la covariance empirique (filtre passe-bas). La méthode semble récente et encore peu appliquée en océanographie (Alvera-Azcarate et al. 2004), et est analogue à l'Analyse en Composantes Principales sur données manquantes (Troyanskaya et al. 2001). Ses avantages sur l'analyse objective classique sont sa grande rapidité (jusqu'à 30 fois), pour des erreurs d'estimation comparables (Alvera-Azcarate et al. 2004). C'est enfin une méthode particulièrement adaptée aux jeux de données grillés (compilation d'images, sorties de modèles...) et des variables "difficiles", comme les images bruitées issues de radiométrie. Pour une cartographie à changement de dimension (1D->2D) comme l'altimétrie, où la hauteur de mer est relativement bien connue en terme de variance spatiale (Le Traon et al. 1997), l'analyse objective semble être plus indiquée.

Les efforts de représentation et de modélisation des fronts océaniques (autres que ceux des grands systèmes comme le Kuroshio ou le Gulf Stream) sont encore marginaux en océanographie: si les modèles les plus avancés résolvent aujourd'hui la méso-échelle ("eddy-resolving OGCMs"), ils se révèlent peu capables de représenter la dynamique des fronts telle qu'observée en imagerie satellite. Leur importance dans la délimitation de structures pourrait par exemple leur valoir une méthode d'assimilation dédiée (e.g. morphisme). **L'imprédictibilité** de nombreux types de fronts rend toutefois difficile leur modélisation déterministe: leur cartographie globale ou régionale reste plus que jamais nécessaire, de même que leur appréhension comme des variables aléatoires spatialisées. C'est en tant qu'**entités écologiques** (i.e. comme habitats pour les organismes pélagiques) que nous continuons ici leur étude.

Les fronts ont ici une double définition, et peuvent être vus à la fois comme interfaces et gradients. La grande non-linéarité et la multiplicité des mécanismes à leur origine leur confère une grande hétérogénéité spatiale et temporelle.

2.2. Les fronts et la production biologique

Un postulat classique en écologie spatiale est celui de l'habitat homogène (i.e. de variance stationnaire et raisonnablement distribuée) ou partiellement homogène (e.g. dans l'étude des "patches" ou des métapopulations, et pour des données grillées en général). L'interface séparant deux milieux homogènes complique la panoplie des mécanismes régissant la dynamique spatiale des individus: ces interfaces peuvent accélérer ou limiter les échanges par diffusion, être infranchissables, réfléchives, léthales, ou même constituer des habitats à part entière (Olson et al. 1994). Cette organisation de l'environnement contribue à structurer les communautés d'organismes, à moduler leur distribution en gradients et ruptures, à définir des **écotones**.

La notion de frontière a donc reçu une attention particulière en écologie terrestre (Gosz 1993, Risser 1995, Chen et al. 1996, Fagan et al. 2003), dans des études illustrant leur rôle dans le transport de matière (Wiens et al. 1985), la délimitation d'habitats (Woodroffe and Ginsberg 1998) ou comme obstacle au mouvement (Bider 1968). L'impact des fronts sur la biologie des espèces marines a lui aussi fait l'objet de nombreux travaux (Fournier et al. 1984, Savidge and Lennon 1987, Brandt 1993, Largier 1993, Takeoka et al. 1993). Leur influence sur les niveaux trophiques primaire et secondaire est ainsi très documentée (le mode advectif étant le premier facteur de variabilité pour ces communautés): les fortes abondance de plancton dans ces zones ont été attribuées aux phénomènes locaux de convergence, ou directement à la production in-situ, par remontée de nutriments dans la zone euphotique du front (Le

Fevre 1986, Martin et al. 2002). La persistance des mécanismes d'enrichissement et de rétention (Franks 1992) conditionne ensuite la diversité spécifique et la production totale du système (*i.e.* selon les constantes de temps des différentes espèces planctoniques présentes – Brandt et Wadley, 1981, Medvinsky et al., 2001). La différence de température de part et d'autre d'un front thermique est aussi souvent suffisante pour affecter les taux de croissance, contribuant à accentuer les gradients biophysiques trans-fronts (Raymont, 1980). Les propriétés géochimiques des deux masses d'eau auront des effets similaires, en fonction de la sensibilité des espèces présentes aux éléments limitants (*e.g.* diatomées contre coccolithophoridés) (Mackas et al. 1991). Les développements phytoplanctoniques dans les zones frontales sont donc régulièrement associés à des assemblages de zooplancton (Dufour and Stretta 1973, Clark et al. 2001).

Par définition, le micronecton est capable de surmonter l'advection et de se déplacer activement dans l'environnement pélagique. Ces organismes restent toutefois affectés par la circulation et la convergence au niveau des fronts lors de leur stade planctonique. La survie, le développement et la croissance des œufs et des larves sont alors directement conditionnés par les conditions environnementales (*e.g.* dispersion, turbulence, température). Un front peut ainsi jouer un rôle de concentration des larves (Bakun 1996), de barrière à l'advection (bordure infranchissable) vers des zones défavorables comme le large (Munk et al. 1995) ou de zone létale (bordure absorbante) pour des niveaux anormaux de turbulence (Cury and Roy 1989). Lorsque les capacités natatoires des individus sont suffisantes, les mécanismes potentiels d'association fronts/biotope deviennent plus nombreux, et leur séparation moins évidente. Les espèces les moins mobiles (*e.g.* myctophidés, euphausidés) peuvent trouver dans les zones frontales des flux convectifs facilitant la migration nyctémérale et le maintien dans la colonne d'eau. Les travaux sur l'association entre fronts et prédateurs pélagiques restent relativement peu nombreux au regard du nombre d'espèces pélagiques et de l'ubiquité des fronts océaniques les thonidés et les cétacés sont le sujet de la plupart des observations. La communauté halieutique japonaise a dénommé le phénomène "loi de Kitahara" (les travaux de Uda, 1935, rapportent déjà une corrélation entre des bancs de *Katsuwonus pelamis* et des zones frontales). Des bancs de thon germon (*thunnus alalunga*) ont de même été signalés dans des systèmes frontaux sur les côtes du Pacifique (Laurs and Lynn 1977, Laurs et al. 1984). On peut noter que de nombreux travaux reliant l'abondance locale d'une espèce à une isoplèthe quelconque révèlent souvent la signature de structures frontales ou au moins convergentes (Kimura et al. 1997, Polovina et al. 2000, Polovina et al. 2001): une certaine plage de température (*e.g.* l'isotherme 17°C pour la zone de transition du Pacifique nord) peut ainsi définir de façon assez précise la position de fronts planétaires. Pour des fronts intermittents ou liés à la géostrophie, la seule définition d'une isotherme ne suffit plus (Ullman and Cornillon 2000)

L'identification des causes de l'agrégation aux fronts est limitée par la collecte de données auxiliaires précises (e.g. composition en espèce et en taille des bancs de proies potentielles présentes). La difficulté de recueil de ces données, et l'imprédictibilité de la plupart des systèmes frontaux ont mené quelques auteurs à considérer des méthodes de simulation (Brandt 1993, Dagorn et al. 2000, Kirby et al. 2000). Ces travaux montrent une approche similaire, mais différent par le mode de représentation ou d'optimisation choisi (i.e. EDP, réseaux de neurones et programmation dynamique stochastique, respectivement). Leurs résultats se basent sur un équilibre dynamique entre contraintes de thermorégulation (i.e. selon le gradient trans-front de température) et contraintes énergétiques (i.e. selon la distribution trans-front de proies potentielles). L'hypothèse thermorégulatrice doit toutefois être modérée: la stratification verticale des océans est en effet à la fois plus prononcée et plus prédictible à l'échelle du poisson (voir e.g. Brill et al., 1999 pour plus de détails). Il reste toutefois avéré que les fronts fournissent un grand nombre d'**habitats** potentiels dans un volume limité, i.e. l'intersection d'un grand nombre de plages restreintes de variables biogéochimiques (Coutant 1977, Brandt 1993).

L'association *Thunnus thynnus*-fronts a été relativement peu rapportée dans la littérature, et rarement de manière quantitative (Maul et al. 1984, Liorzou 1989, Arrizabalaga et al. 2002, Schick et al. 2004). L'article 2 rapporte une telle association dans le Golfe du Lion, inférée à partir de la campagne de survols STROMBOLI et des mesures des radiomètres AVHRR et SeaWiFS. La **résolution** et la **couverture spatiale** des données employées sont deux points marquants de ce travail: en utilisant des données **indépendantes** de toute activité de pêche, à **très haute résolution** (points GPS pour les bancs de *Thunnus thynnus*, et images de radiométrie quasi-quotidiennes, de résolution au sol 1.1 km), on tente d'assurer une couverture objective de la zone, à l'échelle la plus fine possible. La nébulosité reste toutefois l'élément limitant (le repérage s'effectuant à vue et les radiomètres étant mis en difficulté par les zones nuageuses). Une chaîne de traitement L1-L2 a été implémentée sous IDL et SeaDAS afin d'obtenir une cartographie de la teneur en chlorophylle-a. Des techniques d'analyse d'images (partie 1 de ce chapitre) ont ensuite été appliquées pour extraire les fronts thermiques et chlorophylliens pour chaque carte. La **méthodologie statistique** employée constitue un autre point important, car on aborde là le problème de la mise en relation de données géophysiques **continues** (grillées) et d'observations d'entités biologiques **ponctuelles**: ce point sera détaillé et discuté dans la suite du manuscrit. Les principaux résultats de ce travail concernent les **échelles de variabilité** de l'habitat de surface. On confirme en particulier l'importance des fronts dans sa structuration, et leur probable influence sur les proies potentielles de ce prédateur (e.g. petits pélagiques: clupéidés...). Plusieurs mécanismes semblent influencer la distribution des bancs: si la variabilité physique apparaît prépondérante sur une plage restreinte d'échelles (i.e. 10 à 40 km), la petite (<10 km) et grande (> 40 km) échelle reste inexplicée. En particulier, une sur-agrégation (par rapport à une distribution libre aléatoire) a été détectée à petite échelle. La

téledétection (visuelle et radiométrique) permet une couverture confortable de la zone continentale, mais l'absence de données in-situ (la fameuse "réalité terrain") limite les conclusions quant aux variables responsables du phénomène d'agrégation. Les fronts dans le Golfe du Lion ont une dynamique et une extension complexes (ils proviennent e.g. de l'interaction entre les eaux dessalées et chargées en matière organique du panache du Rhône et les eaux du plateau, des upwellings intermittents liés au coups de Mistral, des incursions du courant Liguro-provençal favorisées par les coup d'Est...). Les conclusions fortes ne peuvent concerner au mieux que les échelles spatiales de l'association bancs/fronts, et la forme de cette relation si elle est détectée. L'hypothèse majeure pour expliquer cette association reste donc d'ordre **trophique**, le Golfe du Lion étant une zone de chasse connue (Farrugio 1981, Liorzou 1989).



Figure 2-5. Scène de chasse dans le Golfe du Lion: banc d'anchois poursuivi par des juvéniles de thon rouge (photo: H. Farrugio)

La détection de l'association prédateur/front reste relativement complexe (les phénomènes impliqués étant hautement dynamiques et stochastiques): la représentation à haute résolution (tant spatiale que temporelle) de la variabilité des fronts de surface est nécessaire, en fonction de la dynamique océanographique locale (les modes atmosphériques, géostrophiques, bathymétriques nécessitant une approche à différentes échelles). Les données physiques et biologiques sont toutes deux limitantes dans ce cas (exception faite de la mise en relation avec des structures quasi-pérennes). La couverture spatiale et temporelle des données de capture par pêche est ainsi souvent insuffisante pour étudier l'effet des fronts. Le besoin de leur cartographie rigoureuse et d'une association systématique avec des données de présence/absence d'espèces pélagiques est ainsi régulièrement souligné (Etnoyer et al., 2004, Malakoff, 2004). Malgré cela, les zones frontales sont le sujet d'un nombre limité de programmes dédiés (Olson and Podestá 1986).

En explicitant l'organisation hétérogène de l'habitat, on pose un problème d'**éthoécologie**: comment un organisme pélagique exploite-t-il

son environnement pour s'orienter, exploiter sa productivité, s'y reproduire? Le manque de visibilité à longue distance est caractéristique du milieu océanique (par rapport au milieu terrestre/aérien): on s'attend donc à ce que les regroupements de prédateurs mobiles opèrent à des échelles de temps différentes. On remarquera ainsi les observations d'association oiseaux marins/fronts: un front possède une signature visible de loin pour un animal aérien, capable de repérer des agrégations de congénères ou autres, et de se regrouper rapidement (Kinder et al. 1983). Le problème s'enrichit aussi d'une composante **bioénergétique**, *i.e.* quelle est la productivité (l'espérance du taux de croissance) d'un tel milieu hétérogène pour un prédateur pélagique?

Les fronts composent un habitat dynamique à replacer dans des échelles plus larges: on aborde donc dans la suite le problème **méthodologique** de la détection d'une relation individu/front.

Au terme de cette partie, il apparaît que les fronts sont des zones privilégiées de transfert d'énergie le long des chaînes trophiques. S'ils sont largement étudiés pour leur impact local sur les niveaux primaires et secondaires, leurs rôle précis reste mal connu. Quant aux niveaux trophiques supérieurs, leurs inter-relations semblent hautement stochastiques et l'observation trop souvent anecdotique. L'effet des fronts sur la productivité biologique globale reste donc un champ de recherche encore largement ouvert, propice à l'expérimentation et à la mesure.

2.3. De la relation environnement-ressource

La mise en relation d'une observation avec des variables explicatives potentielles est basée sur le test d'hypothèse: la construction d'une hypothèse nulle est donc nécessaire. Les phénomènes étudiés ayant une dimension spatiale non négligeable (pour la variable biologique ET les variables géophysiques), cette hypothèse nulle doit être spatialisée. Il ne suffit pas de spécifier une distribution complètement aléatoire pour le compartiment écologique: il est nécessaire de postuler un modèle nul reliant ce compartiment aux variables géophysique. Un court détour théorique sur la notion de Distribution Libre Idéale est ici nécessaire.

L'hypothèse nulle. La distribution libre idéale, ou "Ideal Free Distribution" (Fretwell and Lucas 1970), est une pierre angulaire de l'écologie comportementale et constitue encore aujourd'hui un champ actif de recherche (plus de 700 références au Science Citation Index, 1998-2002). Alors que l'IFD a été initialement développée pour décrire la nidification ou la recherche de nourriture chez les oiseaux terrestres, ce

concept a gagné en ampleur et en maturité pour devenir une théorie générale de l'exploration trophique de l'habitat animal et végétal. L'IFD postule que tous les individus possèdent une connaissance complète et égale de la favorabilité de l'environnement, et que chacun est libre de se déplacer d'une zone à l'autre sans contrainte: on montre alors que la densité locale d'individus est fonction (i) du niveau d'interaction entre individus (e.g. compétition/prédation), et (ii) de la favorabilité des conditions ambiantes. C'est donc la résultante de facteurs densité-dépendents et densité-indépendents. La notion de "préférence environnementale" traduit cette colonisation préférentielle de certaines portions de l'habitat. A l'équilibre, une population sujette à l'IFD présente donc un taux de croissance moyen égal au taux de croissance *per capita*, quelle que soit la structure de l'habitat.

"Ho: Distribution Libre Idéale". L'IFD propose un modèle nul de distribution spatiale qui est souvent réfuté **en l'état** (pour s'en convaincre, on notera la grande diversité de taux de croissance réalisés chez des individus de même espèce, de même cohorte). C'est la **dévi**ation ou le **rejet** de l'hypothèse nulle qui informe donc sur les mécanismes régissant la distribution des individus. Une connaissance incomplète de l'environnement tendra par exemple à augmenter leur dispersion et leur densité dans des habitats *a priori* moins favorables (Kennedy et Grey, 1993). Une perception restreinte, l'interférence entre groupes d'individus ou les différences individuelles dans les aptitudes et stratégies à explorer le milieu auront des conséquences similaires (Abrahams, 1986, Houston et McNamara, 1988).

Dans l'hypothèse où un habitat hétérogène génère effectivement une distribution hétérogène des individus, on peut poser les questions suivantes: comment mettre en évidence les déterminants de l'association ressource-environnement? Comment dégager une "préférence environnementale" à partir d'observations dans l'espace? Deux voies sont possibles, entre l'approche **mécaniste** et l'approche **phénoménologique**. L'approche mécaniste, bien que puissante, repose sur la construction d'un modèle incluant des processus de décision, déplacement, capture, assimilation... dont la connaissance objective n'est pas toujours assurée (Fortin and Jacquez 2000). Nous traiterons dans cette partie de l'approche phénoménologique (*i.e.* l'analyse de séries spatio-temporelles).

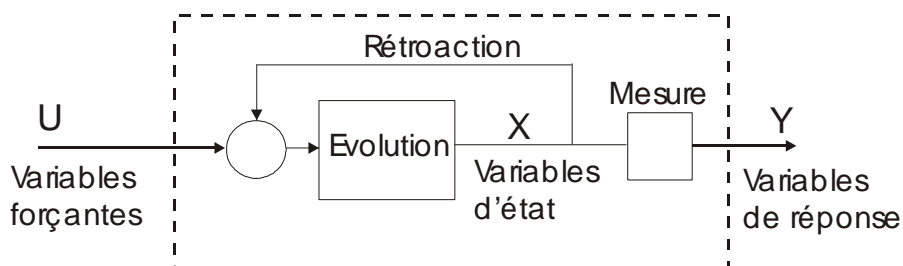


Figure 2-6. Formalisme des systèmes dynamiques, reliant la mesure à des variables forcantes

Deux types d'observation sont en général disponibles: les mesures issues de stations (quel que soit leur arrangement spatial: transect, aléatoire, grillé...) et les mesures dites de distribution ponctuelle (Ripley 1977). Dans les 2 cas, il s'agit de mettre en relation une variable forçante $U(x,t)$ et une observation $Y(x,t)$. Intuitivement, un certain degré d'équilibre entre Y (distribution d'individus) et U (favorabilité de l'environnement) est attendu pour pouvoir tenter de détecter la relation $Y=F(U)$: en particulier, la distribution libre idéale est présupposée dans ce genre d'approche.

L'écueil de l'autocorrélation. Un aspect fondamental des mesures géophysiques et biologiques est leur **autocorrélation**. Ceci est loin de représenter une curiosité statistique: l'autocorrélation décrit comment la connaissance d'une variable à un point donné nous renseigne sur sa valeur en d'autres points de l'espace euclidien. En cela, elle résume en partie la **prédictibilité** d'une variable (Legendre and Legendre 1998). L'autocorrélation d'une série $Y(x,t)$ peut être la signature d'évènements passés (en t) ou voisins (en x): elle informe donc sur les propriétés d'**inertie** du phénomène étudié (*i.e.* sa capacité à amortir et propager une fluctuation). L'autocorrélation peut donc être vue comme un outil d'aide à la détection de pattern: elle reste pourtant souvent perçue comme un effet indésirable qu'il convient de négliger ou, au mieux, corriger (Legendre 1993). Les méthodes statistiques classiques (régression, ANOVA, GLM) souffrent en effet d'une violation de l'hypothèse d'indépendance en analyse spatiale (Legendre and Fortin 1989, Fortin et al. 2002). L'autocorrélation est alors considérée comme un **artefact** (issue d'un processus de régularisation, d'intégration ou d'interpolation dans le temps et l'espace), un effet **indirect** (provoqué par une dépendance à une covariable non mesurée mais autocorrélée dans l'aire d'étude), ou un effet **direct** (correspondant à une réelle dépendance entre densités d'individus proches, *e.g.* un phénomène de gréganisme). De nombreuses autres difficultés techniques existent selon la question et le type de données (Dale et al. 2002).

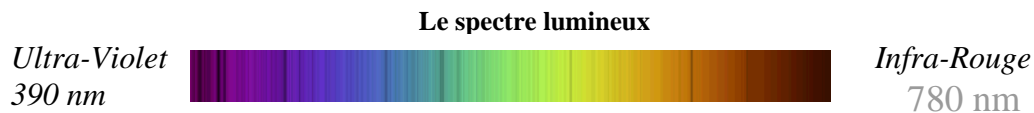
L'analyste fait donc face à plusieurs désagréments: (i) l'autocorrélation entre mesures proches altère la significativité des tests d'une régression ou d'une analyse de variance (Cliff and Ord 1981), (ii) relier $Y(x,t)$ à un environnement autocorrélé $U(x,t)$ fait courir le risque de détecter plusieurs types de relations fonctionnelles possibles, ou des relations inexistantes (Legendre 1993), (iii) les interactions locales (*e.g.* gréganisme) entre individus provoqueront une surestimation de l'intensité de ces relations. Ce dernier point constitue un départ notoire à l'IFD. On montre enfin dans la suite que cette distribution idéale est mise en défaut pour des environnements aléatoires dits "blancs", totalement décorrélés et donc localement non prédictibles: l'absence relative de gradients à grande échelle rend l'environnement beaucoup moins lisible en terme de "gradient de favorabilité" et viole un postulat important de l'IFD (Tyler and Hargrove 1997). Il y a donc un paradoxe certain à tester l'hypothèse d'une association optimale individu-environnement, que celui-ci soit structuré ou non.

De la couleur du spectre d'une série écologique

On rappelle que tout signal temporel peut s'écrire dans le domaine fréquentiel sous la forme de séries de Fourier:

$$X(f) = \int_{-\infty}^{\infty} x(t) \exp(-i2\pi ft) dt \rightarrow X(f) = R(f) + iI(f) = |X(f)| \exp(i\phi(f)),$$

où $e^{-ix} = \cos(x) - i \sin(x)$, et R et I sont les parties réelles et imaginaires de X(f). On appelle **spectre de Fourier** l'amplitude $\Phi(f) = |X(f)|$, qui décrit la répartition de la variance du signal en fonction de ses fréquences. Une représentation en log-log permet alors d'apprécier le ratio hautes fréquences/basses fréquences caractérisant le signal. Lorsque le signal se structure selon une relation $\Phi(f) = 1/f^\beta$ (intégralement ou par morceaux), on parle de **bruit en 1/f**. De nombreuses séries géophysiques et biologiques sont ainsi structurées (Mandelbrot and Wallis 1969, Halley 1996, Ripa and Lundberg 1996, Gisiger 2001). Par analogie au spectre lumineux, on dénomme **bruit blanc** tout signal contenant également toutes les fréquences ($\beta=0$), **bruit rouge** un signal caractérisé par de plus basses fréquences ($\beta=1$), et **bruit bleu** un signal dominé par les hautes fréquences ($\beta=-1$). Sans surprise, le mouvement brownien est lui appelé ... **bruit brun** ($\beta=2$).



Le problème de l'autocorrélation étant largement reconnu en statistique classique (Student 1914), il existe ainsi de nombreuses méthodes pour la détecter et corriger ses effets (Haining 1990; Ramsey et Schafer 1997), comme les tests développés par Geary, Moran, Hubert, ou Mantel (Haining 1990). La simple vérification des résidus permet de détecter l'autocorrélation spatiale (voir e.g. Rathert et al. 1999, Hilborn et Mangel, 1997), et d'entamer la recherche de variables candidates pouvant compléter le modèle. En l'absence de variables forçantes supplémentaires, une tendance spatiale peut être incluse au travers d'un indice géographique (e.g. latitude ou longitude), ou estimée *a priori* et soustraite aux observations pour ne travailler que sur des anomalies (c'est la détendancialisatation ou "prewhitening", bien connue en séries temporelles, Chatfield, 1999). Le sous-échantillonnage de la matrice de mesure (e.g. en considérant un distance inter-stations ou inter-pixels supérieure au rayon de décorrélation) permet de limiter l'autocorrélation, au prix d'un signal aliasé (i.e. les processus opérant à haute fréquence deviennent indétectables). Enfin, le seuil de significativité du test peut être ajusté en estimant le degré de liberté réel du système (e.g. 20 observations corrélées pouvant correspondre à seulement 10 cas informatifs, Haining 1990).

Méthodes de calcul de l'autocorrélation spatiale

Supposons une variable X mesurée en trois points A, B, C (C étant un point éloigné). Si le processus est de moyenne nulle et de variance σ^2 , alors $\text{var}[X(A)-X(B)]=2\sigma^2-2\text{cov}[X(A),X(B)]$ et $\text{var}[X(A)-X(C)]=2\sigma^2-2\text{cov}[X(A),X(C)]$. Si la covariance spatiale est positive et décroît avec l'éloignement, alors $\text{cov}[X(A),X(B)]>\text{cov}[X(A),X(C)]$, et ainsi $\text{var}[X(A)-X(B)]<\text{var}[X(A)-X(C)]$. Tout type de mesure de variance permettrait donc de quantifier l'autocorrélation. Si X est de taille N et de moyenne \bar{X} :

INDICE DE MORAN

C'est l'un des plus anciens indices (Moran, 1950)

$$I = \frac{N \sum_i \sum_j W_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{i,j}) \sum_i (X_i - \bar{X})^2},$$

où X_i la réalisation de X en i, et W_{ij} une matrice de poids (généralement une matrice contiguïté). I varie entre -1 et +1.

INDICE DE GEARY

C'est la version "locale" de l'indice de Moran car il concerne les déviations (Geary, 1954):

$$C = \frac{(N-1) [\sum_i \sum_j W_{ij} (X_i - X_j)^2]}{2 (\sum_i \sum_j W_{ij} (X_i - \bar{X})^2)}, \text{ C variant entre 0 et 2.}$$

C= 1 dénote l'absence de corrélation.

CALCUL CLASSIQUE

$$\text{Corr}(X)_j = \sum_{k=0}^{N-1} (X_{j+k} - \bar{X})(X_k - \bar{X}) \quad (\text{nécessite de connaître la moyenne spatiale})$$

TRANSFORMEE RAPIDE DE FOURIER

$\text{Corr}(X) \Leftrightarrow F(X)F(X)^*$ d'après le théorème de corrélation discrète. F(X) est la transformée de Fourier de X, * est la conjugaison complexe.

SEMI-VARIOGRAMME

$$\hat{\chi}(h) = \frac{1}{2N(h)} \sum_{N(h)} (X_i - X_j)^2 \quad \text{Similaire au C de Geary, pour divers niveaux de proximité.}$$

PROCESSUS PONCTUELS ET INDICE DE RIPLEY

$$\hat{K}(d) = \frac{1}{\lambda} \sum_{i=1}^n \sum_{j=1}^n (\delta_{ij} (\|x_i - x_j\| \leq d)) / n \quad K(d) = \pi \cdot d^2 \text{ pour un processus de Poisson}$$

APPLICATIONS

Test et inférence, krigeage, simulation conditionnelle, calcul du périodogramme...

Les relations environnement-ressource halieutique ont été abordées dans de très nombreux travaux basés sur des Modèles Linéaires Généralisés ou Agrégatifs (GLMs/GAMs). Peu de ces travaux traitent explicitement les effets d'autocorrélation dans les matrices de mesure (Bigelow et al. 1999, Maury et al. 2001, Arrizabalaga et al. 2002), ce qui limite l'inférence à partir des seuils de significativité non corrigés. On pourrait résumer ainsi l'approche commune: la variabilité à grande

longueur d'onde est une tendance qu'il convient de retirer ou d'estimer par des indices géographiques (zonaux ou méridiens), la variabilité à moyen terme (cycles pluri-annuels ou méso-échelle) est vue comme la plus informative, alors que la variabilité locale (haute fréquence) est entachée de bruits de mesure ou de processus (Hilborn and Mangel 1997). En effectuant cette décomposition *a priori* et en négligeant l'autocorrélation de nos données, notre capacité à identifier des patterns importants devient limitée (Fortin and Jacquez 2000). La **décomposition de séries spatio-temporelles** est une discipline sensible, qui profite du développement d'outils généraux, souples et robustes: en particulier, les méthodes de **Monte-Carlo** fournissent une réponse commode au problème du test d'hypothèse en statistique spatiale. Il devient alors possible de traiter explicitement l'autocorrélation intrinsèque à U (variable de forçage supposée) et Y (variable de réponse supposée).

Méthodes de Monte-Carlo. On traite dans cette partie des méthodes de permutation appliquée à l'analyse de séries spatio-temporelles, en se focalisant sur les **méthodes** de randomisation, et sur le **support** de cette randomisation. En effet, l'hypothèse nulle H_0 est souvent construite (en univarié) en se référant à "ce que l'on **observerait** sous hypothèse de distribution aléatoire". Or nous travaillons ici en bivarié, *i.e.* sur la forme de la relation liant la variable forcante U (dite de contrôle) et la variable observée Y . Doit-on alors spécifier un lien entre U et Y puis (i) effectuer les simulations de Monte-Carlo sur la variable dépendante $Y(x,t)$, et fixer les variables indépendantes (forcantes) $U(x,t)$, ou (ii) préserver $Y(x,t)$ et travailler sur $U(x,t)$?

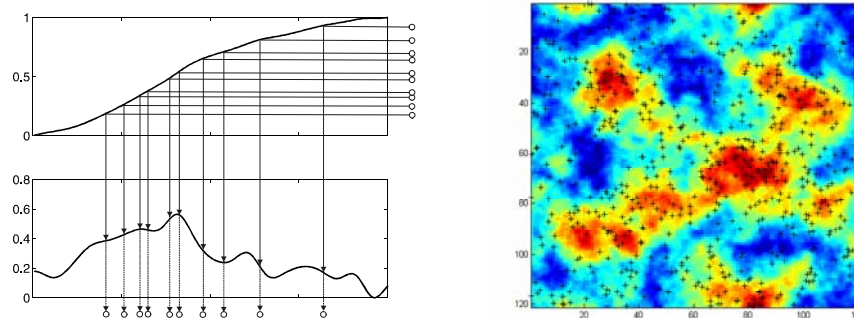
Proposition 1. Dans cette approche, les variables forcantes sont fixées: on applique les simulations de Monte-Carlo sur la variable biologique pour effectuer le test d'hypothèse. Le formalisme des processus ponctuels hétérogènes (Ripley, 1976) aborde naturellement ce problème en considérant des assemblages spatiaux d'entités (d'individus), et en les comparant à une distribution de référence (*e.g.* Poisson, Poisson avec intensité exponentielle, Neymann-Scott). Un **moment statistique d'ordre 2** (*e.g.* K ou L de Ripley) est pour cela défini pour le pattern observé. Un **ensemble** de test est ensuite généré par randomisation relâchée (pour *e.g.* une hypothèse H_0 de processus Poissonien homogène) ou contrainte (pour *e.g.* un processus de Poisson d'intensité hétérogène ou un processus "patchy" de type Neymann-Scott). Cette méthodologie est relativement bien connue en écologie végétale (Ripley 1977), mais est restée apparemment plus confidentielle en halieutique (Kirby et al. Submitted).

Simulation conditionnelle de données ponctuelles

Il est possible, en détenant une c.d.f (fonction de répartition strictement croissante de $[a, b]$ dans $[0, 1]$), de générer tous types de distribution unidimensionnelles ou multidimensionnelles. C'est la **méthode de transformation**, qui permet de projeter un processus stationnaire dans un nouvel espace.

- Se munir d'un générateur aléatoire uniforme $U(0,1)$
- Soit $f(x)$ la distribution cible, de support $[a, b]$. On définit la fonction $g(x) = \int_a^x f(t)dt / \int_a^b f(t)dt$. g^{-1} forme alors une bijection de $[0, 1]$ dans $[a, b]$. En pratique, g peut être quelconque et s'estime alors de façon non-paramétrique (e.g. par polynômes locaux ou splines). $\{x_i\}$ avec $x_i = g^{-1}(N_i)$ fournit ensuite une approximation de $f(x)$ dans $[a, b]$, avec $N_i \sim U(0,1)$.
- L'extension à plusieurs dimensions est immédiate, car l'étape précédente peut être effectuée de façon orthogonale (i.e. en travaillant sur une dimension après l'autre).

EXEMPLES UNI- ET BIDIMENSIONNELS



REMARQUE

L'approche classique en processus ponctuels est la **méthode de rejet**: on génère un processus de Poisson homogène puis on conserve chaque point avec une probabilité $p=1/f$. Cet algorithme est toutefois peu efficace (nombre d'opérations de l'ordre de $2h(b-a)$, où $h=\max(f(x))$). La taille de la population simulée (N) est difficile à contraindre. Des biais significatifs sont à attendre pour des faibles valeurs de N et des surfaces $f(x)$ très hétérogènes.

APPLICATIONS

Méthodes de Monte-Carlo en statistique spatiale, simulation quasi-exacte de champs discrets, accélération d'IBMs, méthodes d'ensemble en assimilation de données, modèles hybrides, résolution d'EDP hautement non-linéaires...

Cette approche est illustrée dans l'article 2, lors du test de l'association entre la distribution des fronts et les bancs de *Thunnus thynnus*. Une première analyse **univariée** (*i.e.* sans lien avec l'environnement) signale (i) le départ d'un processus de Poisson homogène pour l'ensemble des observations journalières 2000-2002, et (ii) un surplus d'agrégation (*i.e.* $K(x) > \pi \cdot x^2$) sur une grande plage d'échelles spatiales (de 5 à 80 km). Cette structuration est confirmée par le variogramme et l'autocorrélation des densités. On peut alors s'interroger sur ses causes: (i) s'agit-il d'un effet de l'hétérogénéité de l'environnement sur la distribution des bancs (cause exogène)?, ou (ii) s'agit-il de la résultante d'interactions entre bancs et groupements de bancs proches (cause endogène)?

La première hypothèse fait écho aux notions de distribution libre idéale ou de préférence environnementale (les bancs devant tendre à coloniser les zones de "favorabilité" supérieure à une valeur moyenne régionale). Or, pour tester cette hypothèse, il est nécessaire de spécifier *a priori* la forme de la relation environnement-densité et ainsi cartographier cette "favorabilité" (approche **bivariée**). Une fois cette surface obtenue, on définit une statistique de test, ou "sonde" (Turchin 2003) afin de tester l'hypothèse H_0 : "les observations sont distribuées aléatoirement conditionnellement à l'environnement $\lambda(x)$ ".

Dans notre cas d'application (article 2), la surface d'intensité $\lambda(x)$ a été spécifiée comme une combinaison linéaire (Besag 1977, Foldager and Pedersen 2002) de la température de surface SST, de la mesure de chlorophylle-a CHLO, et des distances au front thermique ou chlorophyllien le plus proche (DSST, DCHLO). Le schéma de surveillance particulier a été pris en compte en incluant la distance au transect de chaque détection. La statistique de test est le K de Ripley étendu au cas hétérogène pour apprécier l'échelle spatiale résolue par ce modèle: les résultats nous informent que les échelles de 10-40 km apparaissent bien résolues par le modèle complet (incluant les effets du schéma d'observation par transects). Toutefois les échelles inférieures à 10 km signalent un surplus d'agrégation par rapport à notre processus de Poisson. De même, une sur-dispersion est visible pour les échelles plus larges (>40 km). Des réalisations stochastiques correspondant à ce modèle ont été réalisées (en utilisant la méthode classique décrite en encadré). Ces réalisations sont caractérisées par une dispersion supérieure à celle des données à petite échelle, et une densité supérieure à l'échelle de la zone d'étude.

Trois interprétations sont possibles: (i) il existe une variable forçante supplémentaire, non mesurée, opérant à des échelles de l'ordre de 10 km, (ii) les bancs de *Thunnus thynnus*, organisés en clusters, interagissent localement pour générer ces agrégats, (iii) le comportement de chasse/exploration obéit à des mécanismes stochastiques qui brouillent toutes relations directes avec l'environnement local. Le Golfe du Lion constituant un terrain de chasse connu pour les juvéniles, il est probable que ces trois mécanismes interagissent. En d'autres termes,

l'autocorrélation de ce jeu de données a bien été identifiée, mais il n'a pas été possible d'isoler totalement ses causes **directes** et **indirectes** (la haute résolution des données invalidant la possibilité d'un artefact).

Proposition 2. Prenons du recul: en travaillant sur la variable mesurée (Y), il faut admettre que la complexité d' H_0 est limitée par le degré de détail que nous lui accordons (H_0 décrivant généralement un processus trivial ou simpliste que l'on souhaite rejeter). Le rejet successif d'hypothèses de plus en plus complexes est sensé enrichir pas-à-pas notre connaissance, jusqu'à arriver à une description satisfaisante du phénomène (Chamberlain, 1890). Cette approche contient ses propres limitations: la variable de réponse (Y) est l'intégration d'une suite de phénomènes non-linéaires, d'erreurs propagées, de signaux lissés et amplifiés par différents mécanismes dont la connaissance ne peut être intégrale (Jonzen et al., 2002). Dans ces conditions, il apparaît plus judicieux de se pencher sur les **variables de forçage** (U) qui sont en général des variables géophysiques, de dynamique relativement moins complexe que le vivant, et à la fois mieux mesurées, mieux cartographiées, mieux modélisées et mieux prédites.

La reconstruction, l'estimation de données manquantes et la prédiction sont en effet des disciplines bien plus matures en géophysique qu'en écologie: il conviendrait donc d'effectuer les simulations conditionnelles sur U pour tester la forme de $Y=f(U)$. Un rapide tour d'horizon de la littérature nous apprend que cette approche est peu envisagée (Legendre et al. 1990, Sokal et al. 1993). L'emploi de forçages environnementaux dans les modèles d'écologie marine n'est pas systématique, et ils sont rarement remis en cause, remaniés ou perturbés (en dehors d'étapes de rééchantillonnage ou de lissage). Même en océanographie physique, l'utilisation d'ensembles et la perturbation des modèles sont extrêmement récentes. En général, les forçages employés en études "environnement-ressource" sont stockés sous forme de base de données, de matrices de mesure couvrant la période étudiée. Leur taille est souvent insuffisante pour effectuer des répliques *in silico*: l'emploi de modèles de circulation permettrait en théorie de générer à volonté différents états du milieu océanique. En pratique, cette solution est inapplicable car ces outils ont un certain manque de portabilité. Le rééchantillonnage en aveugle (*i.e.* un bootstrap simple) d'un jeu de données détruira la structure temporelle ou spatiale de la variable initiale, et aura peu de sens physique. Une **alternative** existe: le degré de détail et la couverture spatio-temporelle des mesures géophysiques permet d'estimer avec précision leur **spectre de variance** puis de les modéliser dans le domaine **fréquentiel**. Un changement de paradigme est nécessaire: si les méthodes de décomposition en séries de Laplace ou de Fourier préservent la structure fréquentielle du signal, elles détruisent sa structure géographique et temporelle, à la différence des ondelettes (Torrence and Compo 1998). Une vision spectrale est donc privilégiée dans cette approche.

L'objet est alors de générer des séries spatio-temporelles préservant l'autocorrélation du signal originel. Le théorème de la corrélation discrète nous rappelle l'équivalence entre la fonction d'autocorrélation et le spectre de Fourier: préserver l'autocorrélation revient donc à préserver le spectre. Le cœur d'une procédure spectrale de Monte-Carlo serait une décomposition en séries de Fourier (par transformée rapide ou FFT) puis un réarrangement (par bootstrap, ou "shuffling") des phases de chaque série: la transformée de Fourier inverse fournit alors un signal **aléatoire** (toute dynamique ayant été détruite) mais de **même spectre**. En géostatistique, la simulation conditionnelle (Cressie, 1991) a le même objet, mais repose sur la spécification d'un modèle de covariance spatiale (le spectre originel est donc lissé et incomplet). Etant séquentielle, cette méthode est aussi plus lente. Plusieurs variantes peuvent être considérées: implémenté de façon itérative, l'algorithme FFT peut être contraint à conserver aussi la distribution dans le domaine temporel (Venema et al. submitted). L'orthogonalité de la FFT autorise la génération de signaux multi-dimensionnels (2D, 3D, 4D...). Le résultat est donc le suivant: on peut modéliser un phénomène en ne connaissant que son spectre de variance (domaine de Fourier) et sa distribution (domaine temporel). Les phénomènes ainsi générés sont strictement linéaires et n'ont aucune dynamique déterministe. On peut donc les utiliser de façon rigoureuse dans l'étude de la réponse fréquentielle de modèles ou des tests de Monte-Carlo en étude d'habitat.

Quelques aspects importants des statistiques spatiales se dégagent ici: (i) le test de l'IFD par des méthodes classiques semble bien paradoxal, (ii) l'autocorrélation se traite comme un indicateur ET facteur de variabilité, (iii) un processus stochastique peut se caractériser par sa distribution ET son spectre de variance (domaine temporel et domaine fréquentiel), (iv) les tests de Monte-Carlo doivent être privilégiés sur la variable forçante plutôt que sur la variable de réponse.

Simulation conditionnelle de données continues

Trois algorithmes 1D qui peuvent être étendus au cas nD (pour raison d'orthogonalité) afin de générer des signaux linéaires stochastiques de spectre et de distribution données.

BOOTSTRAP DANS LE DOMAINE TEMPOREL

C'est l'algorithme de bootstrap classique: sachant une densité de probabilité $f(x)$, on génère un signal de type "bruit blanc" en rééchantillonnant f , *e.g.* en appliquant la méthode de transformation par splines. La distribution f est conservée, mais le signal obtenu est décorrélé, "blanchi".

BOOTSTRAP DANS LE DOMAINE FREQUENTIEL

Le signal est décomposé en séries discrètes de Fourier (*e.g.* par FFT):

$$X(f) = \int_{-\infty}^{\infty} x(t) \exp(-i2\pi ft) dt \rightarrow X(f) = R(f) + iI(f) = |X(f)| \exp(i\varphi(f)), \quad \text{où } |X(f)| \text{ est}$$
l'amplitude, ou spectre de Fourier, de $x(t)$. Les phases $\varphi(f)$ sont pivotées aléatoirement dans $[0, 2\pi]$: le nouveau jeu de données est ensuite obtenu par transformée de Fourier inverse (iFFT). La distribution originelle en x ou t du signal est détruite (ainsi que toute structure non-linéaire), mais le spectre $|X(f)|$ est préservé (Theiler et al. 1992).

BOOTSTRAP CONTRAINT DANS LE DOMAINE FREQUENTIEL

Cette méthode itérative (Venema et al. submitted) préserve le spectre ET la distribution dans le domaine spatio-temporel: elle s'apparente à une méthode hybride entre la simulation conditionnelle (Cressie 1993) et la randomisation des phases dans le domaine fréquentiel.

PROPOSITION: BOOTSTRAP D'ONDELETTES

Une réalité temporelle ou géographique pourrait être accordée à ces méthodes de simulation, en travaillant sur les phases d'une décomposition en ondelettes (Torrence and Compo 1998). Il devient alors possible de préserver le spectre ET l'hétéroscédasticité. Cette méthode a été récemment introduite pour le cas 1D (Breakspear et al. 2003), et reste à développer pour le cas nD (souvent formalisé différemment dans les méthodes discrètes ou DWT).

APPLICATIONS

Elles sont nombreuses et puissantes, et sont évoquées dans la suite du manuscrit. Il s'agit par exemple de: 1) étendre la réponse fréquentielle (1D) bien connue en théorie des systèmes à des modèles écologiques de toutes dimensions (nD), 2) effectuer des tests de causalité (par simulation de Monte-Carlo) traitant objectivement la structuration spatiale, 3) étudier la réponse d'individus ou de métapopulations à des habitats hétérogènes de diverses caractéristiques...

UN TEST DE TURING...

<http://research.esd.ornl.gov/realizer/>

2.4. Nécessité d'une approche dynamique

Nous sommes maintenant armés d'une batterie d'outils paramétriques et non-paramétriques pour étudier les relations U-Y. Or le problème a été jusque-là abordé sous un angle **statique** (*i.e.* d'une "photographie" de U et Y: la dimension temporelle t n'intervient pas). L'article 2 conclut sur le besoin d'une approche dynamique afin de résoudre les problèmes de distribution spatiale: les techniques de simulation conditionnelle sont ici étendues à trois dimensions (x, y et t) pour illustrer ce besoin.

Diffusion et dispersion dans un milieu aléatoire. Un modèle simple de distribution d'individus est construit sur la base de la chémotaxie telle que définie par Turchin et al. (1997). Les individus évoluent dans un environnement stochastique autocorrélé dans le temps et l'espace (voir encadré précédent). Ils sont sujets à une taxie non destructive (*i.e.* ils n'influent pas sur la favorabilité du milieu) et se déplacent sans consommer d'énergie (selon les hypothèses de l'IFD).

Résolution de l'équation de diffusion par FFT

Soit H un habitat de favorabilité hétérogène $H(x, y, t)$, et $P(x, y, t)$ une densité d'individus se déplaçant dans H selon une diffusion D_2 . On définit arbitrairement H de moyenne 0, et les valeurs positives de H comme favorables aux organismes, *i.e.* ils deviennent sujet à une diffusion de coefficient $D_1 < D_2$. Leur **mobilité** peut être modélisée comme suit (Turchin, Reeve, Cronin et Wilkens, 1997):

$$\mu(H) = \frac{D_2(H+d)}{d+H\frac{D_2}{D_1}} = D_2 - (D_2 - D_1) \frac{H}{H+d\frac{D_2}{D_1}}, \text{ où } d \text{ est un paramètre de saturation reliant la}$$

différence de diffusion à H .

L'équation relation reliant P à H s'écrit alors: $P_t = \nabla^2[\mu(H)P]$

Une solution approchée est fournie par:

$$P(x, y, t) = \frac{\exp(-\nabla^2 \mu(H(x, y, t))t)}{4\pi D_2 t} \left(P_0(x, y) * \exp\left[-\frac{x^2 + y^2}{2D_2 t}\right] \right) \text{ en se rappelant que}$$

$\nabla^2 f = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) f$ et que le Laplacien peut être estimé numériquement. L'opérateur *

dénote ici la convolution, et peut s'effectuer comme une multiplication dans le domaine de Fourier (voir *e.g.* l'estimation d'une densité spatiale de points dans le manuscrit 2).

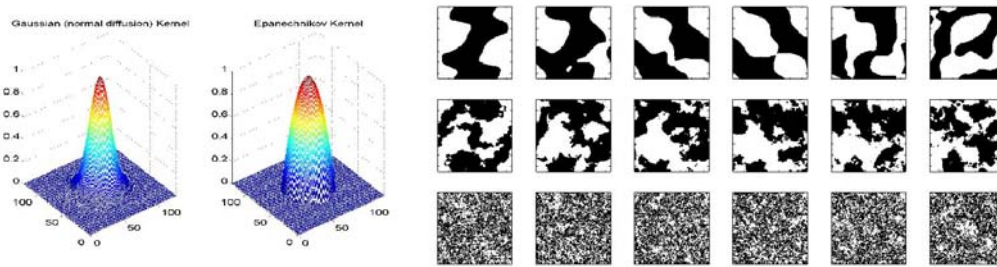


Figure 2-7. A) Noyaux de diffusion de type gaussien (à gauche) et d'Epanechnikov (à droite) utilisés pour la résolution de l'équation de diffusion non-stationnaire. B) Différentes séries temporelles d'espaces pour des pentes spectrales de -2 (bruit brownien), -1 (bruit rouge) et 0 (bruit blanc), de haut en bas.

Des simulations pour divers degrés de structuration de l'environnement montrent que (i) le gradient de favorabilité (variant avec le bruit temporel) est le forçage principal de l'**advection** des individus, (ii) le noyau de **diffusion** employé conditionne largement leur vitesse de redistribution (*i.e.* le noyau gaussien étant non-nul partout, alors que le noyau d'Epanechnikov est de support fini), (iii) l'augmentation du bruit spatial génère un phénomène de **dispersion** (*i.e.* de fission des densités de probabilités).

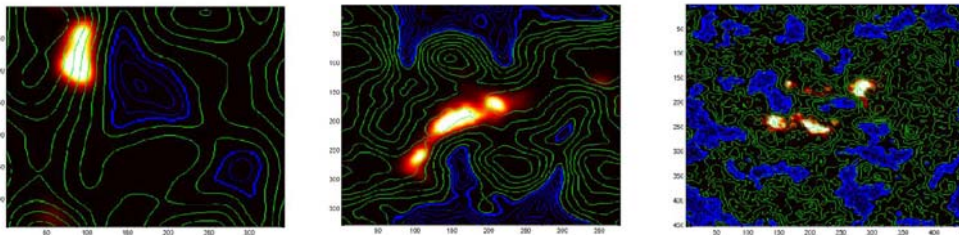


Figure 2-8. Aperçus de l'interaction entre la dynamique agrégative et la variabilité spatiale d'une surface de favorabilité. L'advection est maximale pour des environnements très autocorrélés (gradients de grande échelle). La dispersion (le fractionnement) est maximale pour des environnements de spectre blanc .

Dans cette illustration, la Distribution Libre Idéale apparaît être un modèle nul peu efficace: la présence de hautes fréquences dans l'habitat (bruit blanc ou rose) brouille les gradients à grande longueur d'onde et donc le mécanisme de l'advection à l'échelle de la zone. Le noyau de diffusion employé est insuffisant pour surmonter les gradients locaux et redistribuer la densité P sur toutes les zones favorables: elle est alors sujette à dispersion. En d'autres termes, l'interaction entre la capacité diffusive d'une population et la structuration de l'habitat conditionne sa distribution spatiale à l'instant t .

Ce modèle **déterministe** forcé par un environnement **stochastique** nous confirme le besoin d'une vision dynamique du problème. L'IFD et les outils géostatistiques en général présupposent un équilibre entre la variable mesurée et les variables forçantes: ces méthodes s'avèrent inefficaces si la dynamique étudiée est transitoire, ou si les mécanismes

opèrent à des échelles de temps trop différentes. Tyler et Hargrove (1997) tirent des conclusions similaires d'un modèle stochastique d'IFD, en montrant comment un habitat structuré (*i.e.* positivement autocorrélé) peut occulter une relation habitat/individu.

Nouvelles questions. Nous avons traité dans cette partie de l'approche phénoménologique en statistique spatiale. Un aspect majeur de cette approche est de décrire et comprendre la variabilité des forçages considérés (variables indépendantes) et du signal mesuré (variable dépendante). C'est la condition à la construction d'un test d'hypothèse. La possibilité d'une relation environnement-ressource dynamique révèle alors les limites de ce genre de test: **quantifier la mobilité des individus** dans leur **habitat** devient nécessaire.

Les points-clés de ce chapitre sont: (i) le traitement explicite des fronts comme des discontinuités spatiales, (ii) une approche non-paramétrique du traitement du signal (identification des interfaces pour la détection des fronts, décomposition en séries orthogonales comme alternative à l'analyse objective), (iii) l'emploi des processus ponctuels hétérogènes pour tester la relation banc-front, (iv) une approche spectrale ($1/f\alpha$) de la variabilité environnementale, (v) la proposition de méthodes de Monte-Carlo orthogonales basées sur les séries de Fourier (signaux continus) et la construction de bijections (signaux ponctuels).

3. Du mouvement dans le domaine pélagique

Qui se souvient de la navigation à l'estime ou par les astres? L'estimation de la longitude a été le grand problème de marine du XVIIème siècle. L'invention d'un chronomètre sans dérive l'a résolu. Deux cent ans plus tard, en 1974, une équipe polynésienne redécouvrait un mode de navigation empirique, basé sur les étoiles et les repères naturels à la surface de l'océan. Sur une pirogue traditionnelle, le Hokulea, l'équipage surprit le monde en accomplissant des atterrissages parfaits. Ils démontraient du même coup la possibilité d'une colonisation des îles Hawaïi et de Rapa Nui par une tribu de Bora Bora, et résolvaient une énigme majeure de l'anthropologie du Pacifique. La connaissance astrologique de la latitude et la transmission d'éphémérides chantés en étaient les clés.

*Pour en savoir plus:
An Ocean in Mind
par Kyselka & Nainoa Thompson*

3.1. Reformulation de la problématique

De l'exploration trophique. La question de l'exploration de l'habitat par l'animal est séduisante, particulièrement en biologie marine. Ce milieu difficile à observer et à mesurer pose en effet un problème fondamental, laissant l'homme - animal terrestre - se perdre en conjectures sur les modalités de cette exploration. Dans un habitat globalement peu productif et de faible visibilité, les agrégats d'individus font figure d'anomalies: s'en remettent-ils à la chance pour survivre, soumis à leur environnement, ou ont-ils développés des stratégies cohérentes et efficaces? A moins que notre représentation du milieu ne soit incomplète et trop empreinte d'habitudes "terrestres" (Steele 1985, 1995)? Un banc d'anchois, la migration trans-océanique d'un thon sont autant de témoignages de réponses précises à la variabilité marine.

Peut-on présupposer le type d'information nécessaire pour répondre à ces questions? Un court détour théorique (complétant la Distribution Libre Idéale au chapitre 2.3) peut nous servir de cadre pour la suite du raisonnement. Trente ans de travaux en écologie terrestre ont permis de développer et illustrer de nombreux modèles des déplacements trophiques: le paradigme dominant repose sur un **sélection densité-dépendante** de l'habitat (Kacelnik et al. 1992, Tyler and Hargrove 1997). Ce concept repose donc sur la "favorabilité" apparente du milieu pour un individu donné, qui peut se synthétiser par le **théorème de la valeur marginale** (Charnov 1976). La notion de favorabilité apparente est d'abord valable aux conditions limites, *i.e.* pour des milieux de forte productivité. Dans les autres cas, elle suppose de la part de l'animal une

connaissance de la productivité de son environnement, et la possibilité d'une mesure instantanée de la pression locale de compétition ou prédation. Un "effet mémoire", cognitif ou physiologique (e.g. basé sur les besoins énergétiques), est alors supposé intervenir (Charnov 1976). Dans un environnement de productivité variable (en "patches"), sans compétition où le coût du déplacement varie linéairement avec la distance, cette théorie stipule qu'un individu doit quitter une région lorsque sa productivité devient inférieure à la productivité globale. On sous-entend donc que la **qualité trophique** de l'environnement immédiat peut être évaluée à tout moment (e.g. par des indicateurs comme le spectre de taille des proies) et comparé à la perception de la productivité globale de l'habitat (e.g. sachant un taux de croissance moyen ou une fitness cible). Ceci constitue une vision individu-centrée de la théorie de l'exploration trophique, ou "optimal foraging" (Pyke 1984, Crowder 1985). La formulation continue de ce concept est directement fournie par l'écriture d'un équation logistique, mettant en relation le taux effectif de croissance *per capita* et la capacité de charge du milieu (Mac Call 1990).

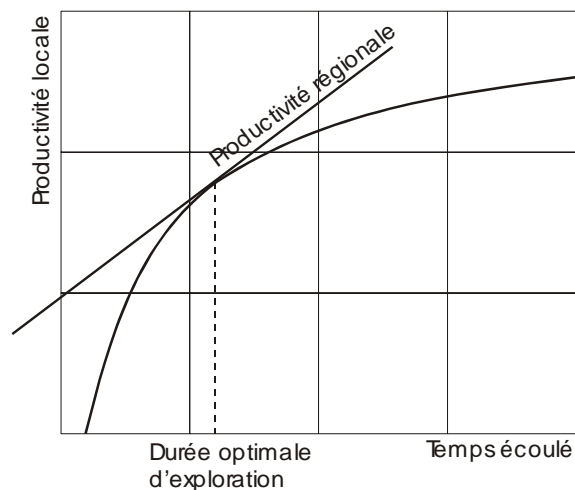


Figure 3-1. Illustration du théorème de la valeur marginale en exploration trophique: lorsque la productivité locale devient inférieure à la productivité régionale, l'individu fait le choix (dans une configuration optimale) de quitter l'endroit exploité.

Un avantage de la formulation individu-centrée de la sélection d'habitat, par rapport à sa version continue, est de reposer principalement sur la dimension temps, plutôt que l'espace. En effet, on attend dans le cadre de cette théorie que les agrégats de proies ne soient pas totalement consommés: la densité caractéristique de renoncement (Giving Up Density – GUD) informe donc sur la **perception** qu'a l'individu idéal de la productivité de son habitat. Le succès de cette théorie en écologie terrestre est dû au fait que l'indicateur GUD est très facile à mesurer et interpréter (Brown et al. 1997). Si le suivi du comportement animal est en temps réel ou enregistré à haute résolution, le temps caractéristique de renoncement (Giving Up Time, GUT) peut aussi être mesuré (Breed et al. 1996). En bref, le **temps** passé à différentes activités renseigne sur les stratégies d'exploration de l'individu et les caractéristiques de son habitat (*i.e.* sa représentation de l'environnement).

Deux modes de déplacement au moins sont donc attendus, correspondant à un mode de recherche "interpatche" et un mode de recherche "intrapatche" (Bovet and Benhamou 1988, Belovsky et al. 1989, Newlands et al. 2004). Ce paradigme fournit, ainsi que la Distribution Libre Idéale, le **modèle d'hypothèse nulle**, qu'il convient de réfuter ou d'améliorer en le confrontant à la réalité terrain. La compétition intra-spécifique, la pression de prédation, la stochasticité de la productivité régionale, et plus généralement l'interaction de termes advectifs et diffusifs sont autant de facteurs qui causeront un départ à ce modèle nul. Néanmoins, ce cadre a l'avantage de poser le problème de l'exploration d'habitat comme un **problème inverse** (Brown 1988), et d'explicitier la perception qu'a le poisson de son environnement (cette approche venant compléter les mesures *in situ* - nécessaires - de toute autre variable géophysique d'intérêt).

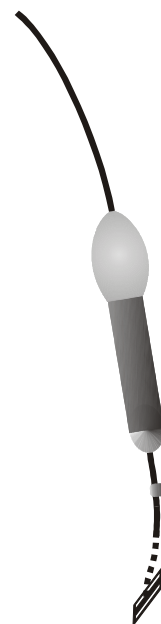
Le paradigme de l'exploration trophique peut ensuite être complété par celui de la thermorégulation: les paramètres de croissance d'un poisson poïkilotherme dépendent en effet largement de la température ambiante (Reynolds and Casterlin 1976). La bioénergétique en éco-éthologie vise alors à unifier ces deux paradigmes, en définissant les gains nets d'énergie fournis par des habitats optimaux, ou l'existence de mesures de conservation d'énergie pour des habitats moins favorables (Wildhaber and Crowder 1990). Enfin, résoudre les relations individus-habitat suppose d'étudier les **modalités** de son exploration, afin d'identifier de possibles mécanismes comme l'orthokinese, la klinokinese, la klinotaxie (DeAngelis and Yeh 1984, Bovet and Benhamou 1988, Alt 1995).

Le **marquage** d'individus en milieu naturel semble définitivement indiqué pour aborder ces questions.

3.2. La marque archive: technologie et promesses

Le suivi des poissons pélagiques a longtemps posé un problème technologique d'observation: une méthode populaire est le suivi par marque acoustique, qui autorise l'étude des mouvements à petite échelle et la mesure de profils océanographiques (e.g. bathythermiques), enrichissant notre connaissance de l'habitat pélagique (Carey 1990). Si la résolution des données fournies est de haute qualité (mesures de l'ordre de la seconde), le coût logistique associé au suivi par navire limite la durée des expérimentations à quelques jours. La **marque archive** fournit une alternative au suivi acoustique, en autorisant des mesures sur plusieurs mois, voire plusieurs années (Gunn et al. 1994).

Technologie. Les marques archives sont de deux types: elles peuvent être **internes**, implantées dans la cavité



péritonéale, ou **externes**, dites "pop-up", ancrées dans la musculature dorsale (Block et al, 1998). Ces instruments sont des enregistreurs dotés d'une mémoire flash non-volatile, associés à des capteurs de luminosité, de profondeur et de température. Les marques internes ont une résolution d'enregistrement de l'ordre de la minute, contre une heure pour les pop-ups, qui nécessitent la recapture du poisson. Après détachement, la flottabilité des pop-ups permet le téléchargement des mesures *via* le réseau Argos, ou d'une position finale seule ("single-point pop-ups").

Les ondes radio ne pouvant pénétrer le milieu aqueux, une mesure de la luminosité autorise une géolocalisation simple en longitude et latitude (basée sur le décalage zénith solaire/midi du chronomètre et la variation de la durée d'ensoleillement avec la latitude). Ces algorithmes (Wilson et al. 1992) sont actuellement implémentés dans des logiciels propriétaires (e.g. Geocontrol v2.01, PatDecoder, Wildlife Computers), et sont mis en défaut par le comportement vertical du poisson (e.g. plongées aux moments critiques comme l'aube ou le crépuscule), la turbidité des eaux de surface, la couverture nuageuse, des erreurs systématiques aux équinoxes (Musyl et al. 2001)... Les marques pop-up souffrent enfin d'erreurs de transmission (20 à 30% de taux d'erreur par message ARGOS de 32 bits) et seraient sujettes à des détachements prématurés, dus à l'enkistement de la pointe d'ancrage (Block and Stevens 2001).

Quelques spécifications

MATERIEL

Le compartiment électronique est enchâssé dans un tube composite de flottabilité légèrement positive, de dimensions 20*180 mm et d'environ 75 g (Wildlife Telemetry). Un câble de nylon et une pointe barbelée en plastique chirurgical permet d'ancrer la marque. La batterie au lithium implémentée actuellement permet d'assurer (en théorie) un an de mesure et la transmission de 10 000 messages de 32 bits sur 7 jours, *via* un émetteur de puissance 0.5 W. Une guillotine mécanique permet le largage automatique de la sonde à la profondeur de 1800 mètres. La taille mémoire associée au microprocesseur est de 2 mégabytes. La précision des capteurs de température et de profondeur est de 1%.

LOGICIEL

Le mode de compression des données peut être précisé, et concerner des histogrammes temps-profondeur, temps-température, des profils température-profondeur, et les profils d'ensoleillement, agrégés de 1 à 24h. La profondeur est mesurée de 0 à 1000m avec une résolution de 0.5m, la température de -40°C à $+60^{\circ}\text{C}$ avec une résolution de 0.05°C . L'irradiance est mesurée à 550nm, de 5.10^{-12} à 5.10^{-2} W.cm^{-2} .

CONSTRUCTEURS

Wildlife Computers, www.wildlifecomputers.com/

Microwave Telemetry, www.microwavetelemetry.com/

Lotek, www.lotek.com/ (marques archives classiques seulement)

Initiatives de marquage. Malgré ces problèmes technologiques et leur coût unitaire, les marques archives délivrent des renseignements précieux et fournis sur l'écologie des thonidés comme les thons rouges (Gunn et al. 1994, Block et al. 1998, Block et al. 1998,

Lutcavage et al. 1999, Block et al. 2001, Boustany et al. 2001, De Metrio et al. 2001, Inagake et al. 2001, De Metrio et al. 2002, Sibert et al. 2003). Les résultats majeurs concernent: (i) leur comportement vertical (échelle 0-500 mètres), et (ii) l'étendue de leurs migrations océaniques (échelle 1000 km). Un grand nombre de ces résultats sont synthétisés par Block et Stevens (2001). Leur interprétation n'est toutefois pas toujours aisée: on note que Lutcavage et al. (1999) signalent des individus potentiellement matures dans l'extension du Gulf Stream au mois de mars, et concluent à la possibilité de zones de reproduction secondaires. La survie après marquage, la délimitation des routes de migration et les interactions possibles avec les pêcheries sont d'autres sujets que les marques archives permettent d'aborder.

Le projet TOPP (Tagging of Ocean's Pelagic Predators), au sein du programme Census of Marine Life illustre l'investissement matériel et financier que peuvent recevoir ces marques. Le marquage d'un grand nombre d'individus (~2500 sur 2003-2005) d'une vingtaine d'espèces cibles a pour objectif de fournir une description de l'habitat océanique "vu par l'animal" (Block et al. 2002). Trois thèmes sont identifiés: (i) comprendre comment les espèces marines explorent et exploitent leur environnement, (ii) discerner des corrélations spatiales entre les mécanismes océaniques et les prédateurs pélagiques, et (iii) identifier les couloirs de migrations et les zones critiques pour ces prédateurs.

Exploitation des marques archives. La discussion du chapitre précédent faisait valoir le besoin d'une vision dynamique, individu-centrée de l'habitat pélagique. Différents types de mesures (archivées ou ancillaires) semblent nécessaires au développement d'une telle vision: (i) des mesures **géographiques** (*i.e.* de localisation des individus), (ii) des mesures **temporelles** (de temps d'exploration, de latence, de piégeage...), (iii) des mesures de variables géophysiques **locales** (*e.g.* température ambiante), (iv) des mesures de variables géophysiques **régionales** (*e.g.* cartographie 3D du champ de température environnant).

L'approche de l'habitat pélagique par les marques archives ne peut alors se résumer à un problème d'ingénierie (*i.e.* de mise au point de capteurs), mais doit nécessairement être duelle (basée sur un test d'hypothèse écologique, spatialement explicite). En pratique, il apparaît que le problème technologique occupe le devant de la scène (Block et al. 2002), alors que l'aspect traitement de données est souvent délaissé malgré son importance (Sims 2003, Boyd et al. 2004).

3.3. Le problème numérique.

L'exploitation de données d'archives peut être abordée selon deux approches complémentaires, *i.e.* phénoménologique ou mécaniste. Une approche purement statistique implique de travailler sur la donnée la plus précise possible: les techniques de traitement du signal classique peuvent être alors employées directement (*e.g.* lissage, interpolation, méthodes de

Fourier - spectre de variance et autocorrélation...). Un différent semble opposer les tenants des modèles les plus parcimonieux (e.g. basés sur le concept d'invariance d'échelle, comme les vols de Levy ou les trajectoires fractales) et les défenseurs de modèles "modaux" (e.g. basés sur la différenciation des modes de déplacement, l'explicitation de paramètres communs et mesurables comme l'azimuth, la vitesse...). On trouvera dans (Turchin 1996) une revue critique de l'emploi de statistiques basées sur l'invariance d'échelle en éco-éthologie. Si ces statistiques (e.g. exposant de Levy, dimension fractale, sinusité) sont utiles dans la comparaison ou la classification de trajectoires (Wiens et al. 1993, Higgins and Strauss 2004), elles sont inaptes à expliquer les mécanismes sous-jacents à l'exploration. En effet l'**invariance d'échelle** nécessaire à leur calcul implique une structuration linéaire en log-log (i.e. une variation continue – donc indiscernable - de comportements). Or la dimension fractale peut ne pas être constante, mais varier de façon continue avec l'échelle d'observation (e.g. dans le cas d'une marche aléatoire corrélée) ou de façon discontinue (e.g. pour une marche aléatoire à plusieurs modes). Ces cas sont effectivement prédits par la théorie de l'exploration trophique (Knopien and Reddinguis 1985), et observés dans la nature (Kareiva and Shigesada 1983, Turchin 1996, Morales et al. 2004): l'approche fractale semble donc contre-indiquée pour la compréhension de leur déterminisme. En postulant l'existence de plusieurs **modes de mouvement** discernables (e.g. une diffusion corrélée) dans un espace 2D ou 3D, on s'autorise la spécification d'*a priori* dans le calcul: on améliore potentiellement la détection de ces modes lors de l'analyse de données. Cette approche s'est révélée fructueuse en écologie terrestre (Alt 1995, Bergman et al. 2000, Morales et al. 2004), et marine (Crowder 1985, Boyd 1996, Brillinger and Stewart 1996, Newlands et al. 2004). En pratique, les modes de variabilité d'une trajectoire peuvent être séparés grâce à différentes techniques: les moments de vitesse angulaire ou linéaire peuvent être décrit par modèle bayésien (Morales et al. 2004), par le calcul de leur co-spectre (Newlands et al. 2004) ou le seuillage de statistiques empiriques (Gutenkunst et al., 2004).

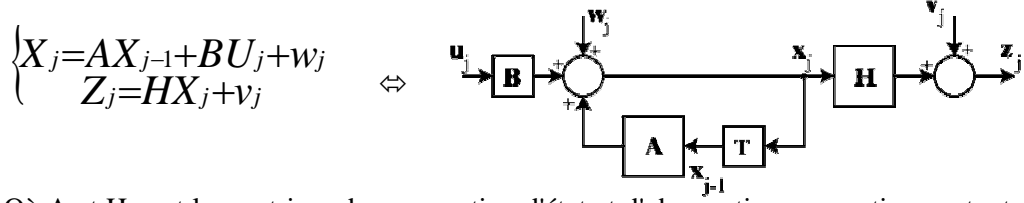
Le succès de ces méthodes (particulièrement en terrestre) est conditionné par la **précision** des données disponibles: les mesures issues de radiotélémetrie ou de suivi acoustique sont propices à ce genre d'étude, malgré leur faible durée (d'autre part, la productivité primaire en terrestre obéit à des échelles de temps beaucoup plus longues qu'en marin, ce qui facilite la détection de relations trophiques). Les marques archives, on l'a vu, présentent des erreurs de géolocalisation importantes, de l'ordre de 1 et 2 degrés en longitude et latitude, respectivement (Musyl et al. 2001, Teo et al. 2004). Ce problème peut être abordé en spécifiant à la fois un modèle de mouvement, et un modèle d'observation. La **théorie des modèles d'espace d'états** ("state-space models") fournit ce formalisme ainsi que des méthodes d'estimation et de filtrage dédiées.

Systèmes dynamiques

De nombreux problèmes d'ingénierie mettent en scène un système devant réagir en fonction de mesures entrée-sortie: la représentation du phénomène sous la forme d'une dynamique cachée, couplée à un processus de mesure ("state-space model") permet alors d'aborder le problème de façon souple. Le système s'écrit comme une équation d'évolution de l'état, complétée par une équation d'observation:

$$\begin{cases} X_j = f(X_{j-1}, \theta_j, w_j) \\ Z_j = g(X_j, \xi_j, v_j) \end{cases}$$

X est le vecteur d'état, et décrit complètement l'état du système à l'instant t. L'évolution de X est définie par la fonction f, de paramètre θ . X est supposé partiellement observable: une mesure est fournie par le processus d'observation g, de paramètre ξ . Des bruits de processus (du e.g. des mécanismes stochastiques ou non décomposables) et de mesure (i.e. n'influençant la dynamique du système) peuvent être introduits sous la forme de w et v. Historiquement, le besoin de calculs en temps réel et les moyens informatiques/analogiques limités ont privilégié la forme **linéarisée** et l'écriture **matricielle** des systèmes dynamiques:



Où A et H sont les matrices de propagation d'état et d'observation, respectivement, et v et w sont les bruits de processus et d'observations définis $p(w) \sim N(0, Q)$ et $p(v) \sim N(0, R)$. F est un biais potentiel de l'observation. U est une variable stochastique (un forçage externe) ou déterminée par l'opérateur (c'est alors un **contrôle**). Dans le diagramme de bloc ci-dessus, l'opérateur T désigne le délai unitaire, i.e. $T(X_t) = X_{t+1}$. L'écriture d'un système sous cette forme n'est pas anodine, et obéit à un certain nombre de règles pour être "bien posé". En particulier, l'observabilité de la dynamique (i.e. la possibilité d'inférer X à partir de Z), sa contrôlabilité (i.e. définir U pour obtenir un certain état cible) permettent d'obtenir la forme dite canonique - inversible et causale.

Une telle notation a des applications nombreuses et puissantes, telles que le **filtrage** (estimation en temps réel de l'état x_j sachant une série passée d'observations $z_{1:j}$), le **lissage** (estimation *a posteriori* de l'état x_j sachant une série complète $z_{1:T}$), l'**estimation de paramètres** (estimation de θ sachant $z_{1:j}$), la **prédiction/simulation** (propagation de x_j), la **réponse fréquentielle** (calcul de la fonction de transfert)...

Dans notre cas d'application (i.e. la localisation d'une marque en deux dimensions), on définit $X_j = \begin{pmatrix} lon \\ lat \end{pmatrix}$ comme le vecteur d'état à estimer. Le système peut alors s'écrire sous la forme suivante (Sibert et al. 2003):

$$B = \begin{pmatrix} u\Delta t \\ v\Delta t \end{pmatrix} \quad A = \begin{pmatrix} 2D\Delta t & 0 \\ 0 & 2D\Delta t \end{pmatrix} \quad H_j^{-1} = \begin{pmatrix} 111.\cos(\hat{X}_{j,2}/111) & 0 \\ 0 & 111 \end{pmatrix} \quad F = \begin{pmatrix} b_x \\ b_y \end{pmatrix}$$

B définissant le biais dans l'advection du mobile (i.e. le processus advectif), A étant analogue à un processus de diffusion classique de coefficient D, H définissant le changement de coordonnées kilométrique-polaire (111 kilomètres étant la distance couverte par 1 degré à l'équateur), et F définissant les biais du processus d'observation:

$p(v) \sim N(F, R)$. On remarque que Sibert et al. (2003) définissent $B.U_j$ comme étant un terme d'advection, alors que le formalisme classique d'un modèle dynamique (dans sa version "contrôlée") pose U comme la variable de commande. Cela fait peu de différence pour la suite de l'exposé: il nous suffit de poser $U_j = Id$ (Id étant la matrice identité).

Paramètres et états, quelle différence?

Une remarque s'impose ici sur la distinction entre **paramètres** et **états**. Lorsque le modèle est stationnaire, les paramètres sont fixés et seul l'état est estimé. Lorsque le modèle est non-stationnaire (*i.e.* les paramètres évoluent dans le temps – ou l'espace), les seules différences entre θ et X concernent la forme de leur dynamique. L'état X est ainsi supposé suivre la dynamique imposée par f , alors que θ , en sa qualité de paramètre, aura une évolution du type:

- Marche aléatoire de variance constante dans le cas d'une **dérive**,
- Probabiliste binomial dans le cas de **changements de régime**,

Intuitivement, si θ obéissait à la même dynamique que l'état X (ou subissait des évolutions de même fréquence), il n'y aurait plus lieu de le définir comme paramètre. Ce serait ... un état.

Le filtre de Kalman. Lorsque le problème porte sur l'estimation du vecteur d'état X , le filtre dit "de Kalman" fournit le meilleur estimateur si la variance des bruits de processus et de mesure sont connus. Si ces bruits sont gaussiens, le filtre de Kalman est optimal: on le dénomme alors BLUE (Best Linear Unbiased Estimator), comme l'estimateur de krigeage. Son mode de fonctionnement est basé sur une pondération entre une information donnée *a priori* (*i.e.* l'estimation de l'état au temps $t-1$) et une nouvelle observation, procurant une **innovation**, au temps t . Kalman (1960) a montré que cette pondération pouvait prendre la forme d'un **gain** K , en cherchant à minimiser la trace de la matrice de covariance du vecteur d'état. Un diagramme de bloc permet d'appréhender la séquence d'assimilation d'une innovation dans le filtre de Kalman:

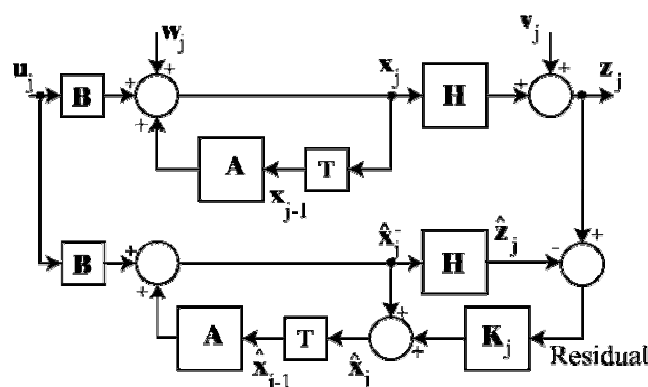


Figure 3-2. Diagramme de blocs du filtre de Kalman couplé à la dynamique réelle.

Dans cette représentation, le modèle de propagation d'état est supposé connu: le même modèle est donc utilisé pour la génération de

l'état réel (inconnu) x et de l'état prédit (\hat{x}). En pratique et en sciences naturelles, c'est très rarement le cas: on doit se contenter d'un modèle approchant la dynamique réelle.

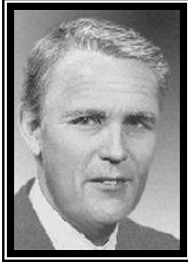
Les équations de Kalman

<p>Prédiction de l'état au temps j</p> <p>Propagation de la covariance associée</p> <p>Calcul du gain de Kalman au temps j</p> <p>Correction de l'état sachant K et l'observation z_j</p> <p>Mise à jour et réduction de la covariance</p>	$\left\{ \begin{aligned} \hat{x}'_j &= A\hat{x}'_{j-1} + Bu_j \\ P'_j &= AP_{j-1}A^T + Q \\ K_j &= \frac{P_j H^T}{HP_j H^T + R} \\ \hat{x}_j &= \hat{x}'_j + K_j(z_j - H\hat{x}'_j) \\ P_j &= (I - K_j H)P'_j \end{aligned} \right.$
---	--

Comme pour l'estimateur de krigeage, la formulation analytique du gain (Kalman 1960) s'obtient en cherchant à minimiser l'erreur de prédiction:

$$P_j = E\{e'_j e'^T_j\} = E\{(x_j - \hat{x}'_j)(x_j - \hat{x}'_j)^T\}$$

$$\frac{\partial P_j}{\partial K_j} = \frac{\partial E\{(x_j - \hat{x}'_j)(x_j - \hat{x}'_j)^T\}}{\partial K_j} = 0$$



Rudolph Emil Kalman

On remarquera que le filtre de Kalman est une généralisation des moindres carrés récursifs, qui considèrent un système sans bruits de processus ni signaux U en entrée (Roweis and Ghahramani 1999). Etant codé sous forme matricielle, un tel filtre est facile à implémenter dans la plupart des langages de programmation scientifique (e.g. R, S+, Matlab). Il est relativement peu gourmand en ressources informatiques puisqu'il ne nécessite pas le stockage complet des schémas passés états/erreurs (celui-ci se faisant implicitement par une mise à jour de la matrice de covariance P). Il est enfin possible, une fois les calculs achevés dans le sens direct, de les effectuer de façon rétrograde (i.e. en utilisant l'information future dans l'estimation de l'état t). Le filtre devient alors le **lisseur** de Kalman, fondamentalement équivalent à l'opérateur de krigeage (Bertino et al. 2002).

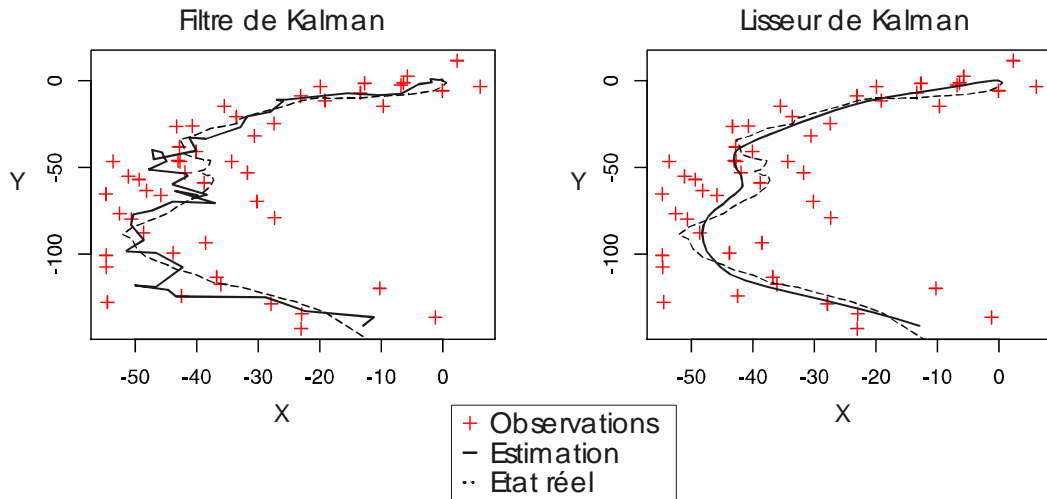


Figure 3-3. Comparaison du filtre (à gauche) et du lisseur de Kalman (à droite), sur un cas contrôlé

De nombreuses **variations** sont possibles. On peut en particulier effectuer une estimation **duelle** de l'état ET des paramètres. Le filtre de Kalman est souvent perçu comme une méthode d'assimilation en temps réel, ou « on-line ». Pourtant l'hypothèse de linéarité et d'erreurs gaussiennes autorise de l'utiliser dans un cadre d'estimation **globale**, car la vraisemblance d'une suite d'itérations s'écrit de façon exacte (Harvey 1989, Roweis and Ghahramani 1999).

$$L = -\frac{1}{2} \sum_j \ln |P_j| - \frac{1}{2} \sum_j (z_j - H\hat{x}'_j)^T P_j^{-1} (z_j - H\hat{x}'_j),$$

où l'on rappelle que P est la matrice de covariance associée à l'état prédit et z sont les observations. On s'éloigne ici de la philosophie originelle du filtre, mais ce détournement permet l'estimation e.g. des matrices de covariance d'erreurs Q et R, rarement connues en réalité. Employé ainsi, un filtre de Kalman s'apparente aux méthodes dites de « Markov Chain Monte Carlo » en séries temporelles, car l'optimisation des paramètres s'effectue de façon **globale** : les équations de Kalman servent alors seulement à écrire de façon compacte la vraisemblance (Kohn and Ansley 1986). C'est cette approche qui a été retenue dans les travaux de (Sibert et al. 2003), combinée à une méthode de gradient pour assurer l'estimation globale du jeu de paramètres d'erreur (i.e. en menant un grand nombre de fois la série complète d'itérations). L'estimation conjointe des erreurs de processus et d'observations pose alors un problème, en particulier lorsqu'elles sont d'amplitudes comparables (voir figure pour l'emploi récursif d'un algorithme génétique).

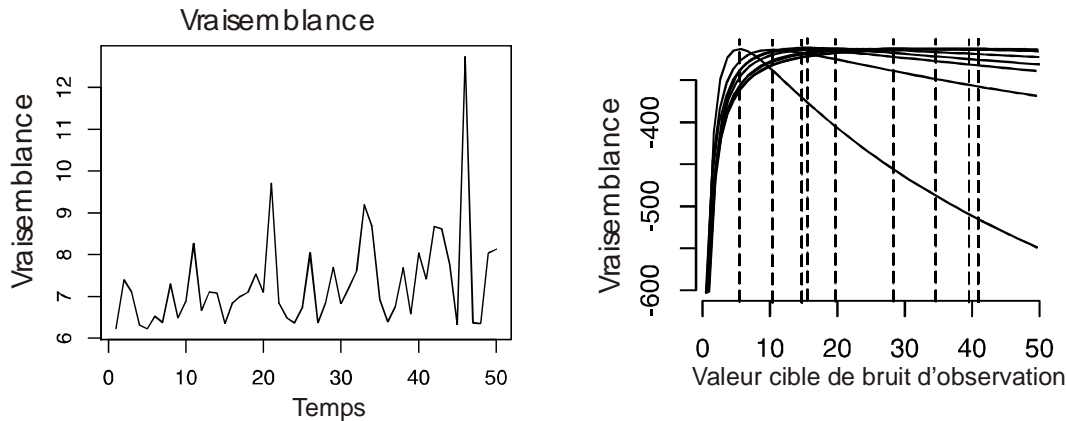


Figure 3-4. Evolution de la vraisemblance avec le temps pour le cas contrôlé précédent (à gauche), et fonction d'optimisation de la vraisemblance d'un paramètre (bruit d'observation) pour diverses valeurs cibles (à droite).

Ce même problème d'estimation des covariances d'erreur peut être abordé **localement** avec une version **adaptative** du filtre de Kalman (Dempster et al. 1977). Un algorithme de Maximisation de l'Espérance est employé séquentiellement (*i.e.* à chaque pas de temps), pour assurer la mise à jour du vecteur d'état ET des matrices de covariance d'erreur. Etat et paramètres sont alors estimés conjointement : la distinction entre les deux notations (X et θ) devient alors moins nette. Intuitivement, un paramètre est supposé être moins variant qu'une variable d'état, ou au moins soumis à une dynamique plus élémentaire (*e.g.* un dérive, ou des variations discrètes).

Le filtre de Kalman Etendu. Bien que séduisant et simple, le filtre de Kalman classique trouve ses limites dans les dynamiques non-linéaires et les erreurs non-gaussiennes. Une version sub-optimale peut alors être obtenue par linéarisation avec un opérateur tangent, le Jacobien (Lewis 1986). Les matrices A et H s'écrivent alors comme les dérivées partielles de f et g en x , et W et V sont les dérivées partielles de f et g en w et v :

$$A = \frac{\partial f}{\partial x} \quad H = \frac{\partial g}{\partial x} \quad W = \frac{\partial f}{\partial w} \quad V = \frac{\partial g}{\partial v}$$

Les équations classiques de Kalman s'appliquent sans modification, et permettent l'estimation sub-optimale de X sous hypothèse d'erreurs quasi-gaussiennes et de dynamique quasi-linéaire : la vraisemblance d'un jeu de données peut là encore être calculée de façon efficace. En pratique, on peut décomposer f et g en séries de Taylor ou utiliser des méthodes d'approximation numérique du jacobien (*e.g.* numjac.m pour Matlab ou jac.r de la librairie R nlmODE).

L'EKF fournit une approximation du processus d'estimation : des dynamiques hautement non-linéaires le mettront en défaut. D'autre part, il sera vraisemblablement lent sur des problèmes de grande dimension (l'approximation d'un plan tangent devenant de plus en plus complexe

avec la dimension). L'étape de calcul du Jacobien à chaque état et pour chaque erreur nominale étant obligatoire, il est préférable de conserver l'EKF pour des problèmes de faible dimension. Une alternative est de se tourner vers des versions réduites, *i.e.* où la dynamique est projetée sur un nombre d'axes réduits correspondant à des fonctions empiriques (EOF, SVD), ou paramétriques (Fourier). Ces versions du filtre de Kalman sont le SEEK (Singular Evolutive Extended Kalman Filter), SIEK (*idem* mais interpolé), ROEKF (Reduced Order Extended Kalman Filter). La simulation d'**ensemble** fournit une seconde alternative.

Le filtre de Kalman d'Ensemble. Le rapprochement entre l'assimilation de données « matricielle » et les simulations de Monte-Carlo ont donné lieu au filtre de Kalman dit « d'Ensemble » (EnsKF). Il a d'ailleurs été introduit pour la première fois sous le nom de « méthode de Monte-Carlo » par (Evensen 1994) dans le Journal of Geophysical Research. Le cœur de la méthode est la simulation conditionnelle d'un nombre limité d'états (« l'ensemble ») reflétant l'erreur associée à l'estimation d'un état au temps t . Cet ensemble est propagé par f (f étant la dynamique non-linéaire originelle) : la prédiction $X_{t|t-1}$ et la covariance d'erreur associée $P_{t|t-1}$ sont ensuite approchées par la moyenne et la covariance d'ensemble (on parle de **troncature gaussienne**). Les équations de Kalman s'appliquent ensuite de façon classique. Un lisseur d'ensemble peut en particulier être implémenté (Evensen and van Leeuwen 2000).

L'EnsKF a été appliqué récemment dans de nombreux problèmes d'assimilation de données (Evensen 2003), de résolution bayésienne et d'analyse de séries temporelles (Gronnevik and Evensen 2001). Son intérêt a été démontré dans l'analyse de séries temporelles non-stationnaires, marquées par des changements de régime ou des dynamiques chaotiques comme le système de Lorenz (Pham 2001). Toutefois l'EnsKF repose sur une troncature gaussienne : si elle est en pratique très permissive, la distribution gaussienne ne permet pas d'approcher des distributions multi-modales. Or la surface d'état peut être caractérisée par une variabilité quasi-cyclique, voire présenter des sauts discrets. Ces discontinuités, courantes en océanographie physique (*e.g.* fronts, tourbillons méso-échelle) ne peuvent être approchées par un modèle gaussien.

Le filtre Particulaire. Lorsque la troncature gaussienne n'est plus justifiable, il est possible d'étendre le filtre de Kalman d'Ensemble à une version généralisée, dite Particulaire (Doucet and Godsill 1998). On l'appelle aussi filtre SIR (pour « Sampling-Importance-Resampling ») ou filtre bootstrap, ou filtre de Monte-Carlo (Dahia 2005).

Aucune hypothèse sur la forme du modèle n'est posée : des densités de probabilités multi-modales peuvent être approchées de cette manière. L'estimation duelle des états/paramètres se fait simplement par **augmentation de l'état**, en incluant θ dans le vecteur X , et en attribuant à θ une dynamique contrainte (*e.g.* une marche aléatoire de variance

constante ou convergente en $1/t$, ou une loi de type binomial) pour estimer e.g. une dérive, des valeurs constantes ou des changements de régime.

Principe

La puissance de calcul des ordinateurs individuels a permis, dès les années 90, l'emploi de méthodes de Monte Carlo pour résoudre certains problèmes non-linéaires d'estimation et de simulation. Contrairement aux méthodes de maillage, ces méthodes sont basées sur la loi des grands nombres et sont donc moins sensibles à la dimension d'espace d'état (elles posent toutefois d'autres problèmes).

On peut affirmer que les méthodes particulières sont une version séquentielle des méthodes de Monte-Carlo (Dahia 2005). Introduites et développées pour des problèmes d'ingénierie (Gordon et al. 1993, Doucet and Godsill 1998), ces méthodes proposent de représenter la loi conditionnelle de l'état par un nombre fini de masses de Dirac pondérées δ .

$$p(X_t|Y_t) \approx 1/N \sum_{i=1}^N w_t^{(i)} \delta(X_t - X_t^{(i)}), i=\{1, \dots, N\}$$

Cet ensemble de points, ou « particules » est associé à un ensemble de coefficients de pondération (les poids) mesurant leur vraisemblance (*i.e.* sachant une observation). Pour éviter la dégénérescence de l'ensemble (la réduction du nuage à un point), une nouvelle étape de rééchantillonnage est introduite afin de dupliquer les particules de poids importants : c'est le « Sampling-Importance-Resampling ».

Le Filtre Particulaire en pratique

Au préalable il faut :

- avoir à disposition une loi de propagation d'état $p(x_k|x_{k-1})$

- pouvoir évaluer en tout point de l'espace d'état une fonction de vraisemblance $p(y_k|x_k)$

Soient $f(x_t, \theta_t)$ et $g(x_t, \theta_t)$ les fonctions d'évolution et d'observation d'un système dynamique, $p(X_0)$ et $p(\theta_0)$ les densités de probabilité initiales de l'état X et des paramètres, $p_{v,t}$ and $p_{r,t}$ les densités de probabilités des bruits de processus et d'observation. Le filtre simulera N particules, soumises à des bruits de rééchantillonnages $p_{r,t,x}$ and $p_{r,t,\theta}$ (contribuant à limiter la dégénérescence).

Initialisation. N particules sont générées en échantillonnant $X_0^{(i)} \sim p(X_0)$, $i=\{1, \dots, N\}$ (voir e.g. la méthode de transformation introduite en §2.3). Soit $w_0^{(i)} := 1/N$ le poids de chaque particule.

Propagation. Pour chaque particule $i=\{1, \dots, N\}$, on échantillonne le bruit de processus $v_t^{X,(i)} \sim p_{v,X}$, le bruit d'évolution des paramètres $v_t^{\theta,(i)} \sim p_{v,\theta}$, et les bruits accessoires $r_t^{X,(i)} \sim p_{r,X}$, $r_t^{\theta,(i)} \sim p_{r,\theta}$ (ces bruits permettent en théorie d'améliorer l'estimation dans le dynamique trop complexe). Les nouveaux vecteurs d'état et de paramètres s'écrivent :

$$\begin{aligned} X_{t+1|t}^{(i)} &= f(X_t^{(i)}, \theta_t^{(i)}) + v_t^{X,(i)} + r_t^{X,(i)} \\ \theta_{t+1|t}^{(i)} &= \theta_t^{(i)} + v_t^{\theta,(i)} + r_t^{\theta,(i)} \end{aligned}$$

Assimilation. Chaque particule voit son poids mis à sa jour sachant sa vraisemblance $i=\{1, \dots, N\}$:

$$\begin{aligned} w_t^{(i)} &= w_{t-1}^{(i)} p(Y_t|X_t^{(i)}) \\ w_t^{(i)} &= w_{t-1}^{(i)} p_{e,t}(Y_t - g(x_t^{(i)}, \theta_t^{(i)})) \\ w_t^{(i)} &= w_t^{(i)} / \sum_l w_t^{(i)} \end{aligned}$$

Rééchantillonnage. L'étape de « Sampling Importance Re-sampling » (Rubin 1987) permet d'éviter la dégénérescence. Un bootstrap est appliqué à l'ensemble $\{X_t^{(i)}(\theta_t^{(i)}, w_t^{(i)})\}$ où $w_t^{(i)}$ sont alors les probabilités des particules. Cette nouvelle distribution fournit une approximation de $p(Y_t|X_t)$.

On notera que le filtre Particulaire et ses nombreuses améliorations sont encore le sujet de recherches actives (Dahia 2005). Si son application à des problèmes fortement non-linéaires a fourni des résultats encourageants (Doucet and Godsill 1998, Fearnhead 1998, Bolviken and Storvik 2001, Papavasiliou 2002), il reste encore une technique expérimentale dont la robustesse est largement dépendante de l'expérience de l'utilisateur.

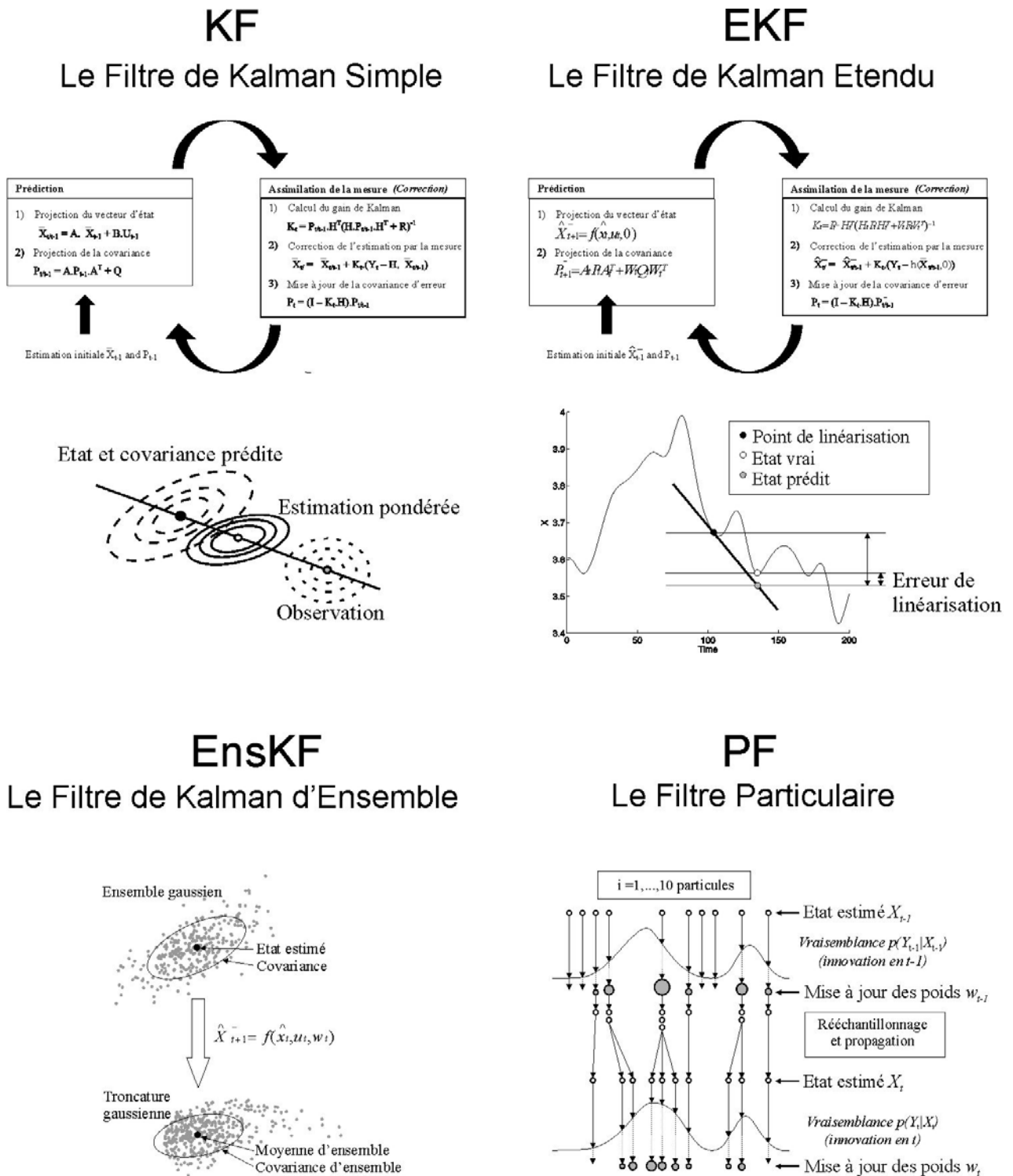


Figure 3-5. Quatre variations du filtre de Kalman et leur fonctionnement schématique

Doit-on conserver la complexité originelle du problème et utiliser une méthode d'ensemble (filtre Bayésien, filtre SIR ou particulière, EnsKF)? Ou réduire cette complexité au préalable (EKF, SEEK, ROKF) et travailler analytiquement? Tout problème d'estimation en séries temporelles semble pouvoir être résolu de façon « brutale » par MCMC et un échantillonneur de Gibbs. Néanmoins la solution « compacte » ne doit pas être négligée, i.e. intégrer le problème sous une forme appropriée de Kalman (KF, EKF, EnsKF) et de tirer parti de la vraisemblance ainsi fournie (Meyer and Christensen 2001).

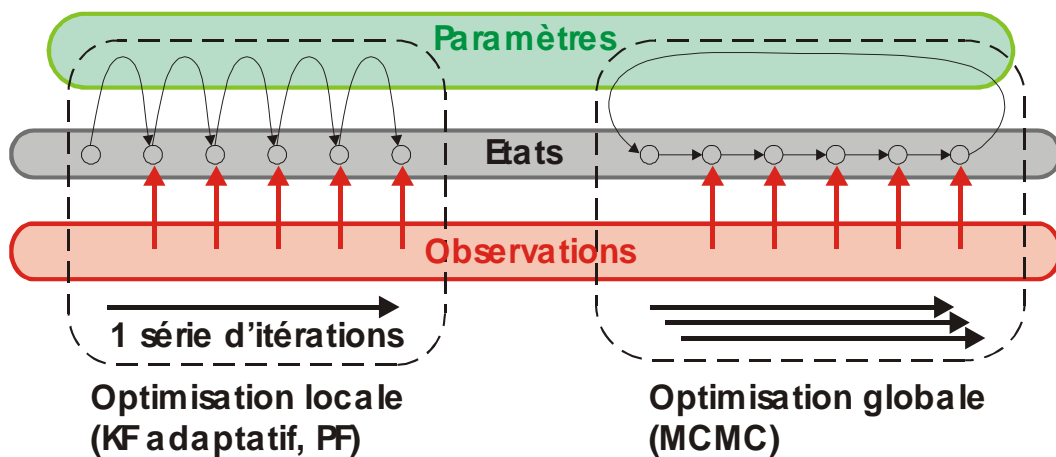


Figure 3-6. Les deux modes principaux d'optimisation (adaptatif ou global) employés en filtrage

3.4. Evaluation de la méthode particulière

Implémentation. Notre problème pourrait donc s'écrire avec un vecteur d'état $X = \{\text{longitude, latitude, profondeur, température, azimuth, vitesse, mode}\}$, où « mode » se réfère au mode de déplacement (e.g. diffusion simple ou corrélée). Longitude et latitude sont corrélées à l'azimuth et à la vitesse du poisson, et donc au mode diffusif auquel il est soumis. La température ambiante est une fonction complexe de la longitude, latitude et profondeur d'évolution. Les observations de géolocalisation sont, on l'a vu, bruitées, incomplètes, biaisées et sous-échantillonnées. Le schéma de biais est en particulier inconnu *a priori*. La profondeur et la température d'évolution sont relativement bien connues (mesure directe), mais le champ de température régional pose le problème de sa représentation (choix entre une mesure de surface par télédétection ou la reconstruction 3D par interpolation de profils ou modèles de circulation). Enfin les vitesses linéaire et angulaire dépendent étroitement d'une bonne estimation des géolocalisations (i.e. une haute fréquence d'estimation ET de faibles erreurs associées).

Présenté ainsi, le problème semble loin d'être simple, même s'il est de relativement faible dimension. Le modèle de déplacement *supposé* étant non-linéaire (de même que le champ de température océanique), les méthodes d'ensemble (EnskF et PF) ont été d'emblée envisagées. Afin d'évaluer objectivement les sources d'erreur et les difficultés d'estimation, différents filtres ont été implémentés sous Matlab et testés sur un jeu de données simulées. **L'article 3** décrit ces étapes et les résultats associés. Deux questions ont été formulées et étudiées :

- L'emploi de mesures de température peut-il améliorer l'estimation des positions (e.g. correction des biais) ?
- La précision de cette géolocalisation permet-elle d'inférer des changements de comportements du poisson marqué ?

C, C++, FORTRAN, MATLAB, OCTAVE, IDL, S+, R, BUGS?

Une grande partie de l'investigation scientifique se faisant derrière un clavier, la question du langage de programmation est inévitable. Les langages dits « de bas niveau » (C, FORTRAN) sont parfaitement indiqués pour les tâches répétitives et intensives en boucles. Les langages interprétés et possédant une structure matricielle sont à privilégier pour la construction d'algorithmes et toutes les tâches pouvant être réduite à des systèmes linéaires. Leur lenteur limite malgré tout leur usage opérationnel.

R et MATLAB sont deux environnements conviviaux : le premier domine par ses fonctionnalités statistiques, et permet le traitement de jeux de données de taille raisonnable. MATLAB incorpore une structure matricielle quasiment sans limites de stockage et de nombreuses fonctions d'approximation et de linéarisation : ce logiciel autorise des applications de grande taille (traitement d'images...) sans perdre la souplesse de la ligne de commande.

La solution d'une résolution MCMC/bayésienne sous WinBUGS (Spiegelhalter et al. 1996) a été écartée car notre cas nécessite l'appel de nombreux fichiers NetCDF et le stockage de matrices de grande taille (fonctionnalités encore non autorisées dans la version actuelle de WinBUGS).

Test. Pour cette étude, une trajectoire fictive a d'abord été simulée en utilisant une diffusion corrélée à deux modes (Bovet and Benhamou 1988, Krakauer and Rodriguez-Girones 1995, Newlands et al. 2004). On se restreint ici au cas bidimensionnel, pour un pas de temps d'une heure. Pour chacun des deux modes, une distribution de vitesse de type Weibull a été construite, ainsi qu'une distribution de vitesse angulaire de type von Mises. Ces distributions paramétriques semblent en effet bien correspondre aux observations de terrain (Lutcavage et al. 2000, Newlands et al. 2004). L'alternance entre modes a été forcée de manière arbitraire, selon une loi binomiale pour la première partie de la trajectoire (de Gibraltar à Capbreton), et forcé en mode « directionnel » pour la seconde. Une série temporelle de température a été construite en échantillonnant une sortie unique du modèle Mercator (Bahurel et al. 2002). Différents schémas de biais/erreurs de localisation ont ensuite été appliqués à la trajectoire sous-échantillonnée à 24 heures : on simule ainsi le processus de géolocalisation, la possibilité d'erreurs corrélées, de

dégradation et de fouling du capteur (Musyl et al. 2001, Sibert et al. 2003). En particulier, des erreurs de l'ordre de 2 degrés en latitude et des biais variables ont été incorporés. Pour tester l'impact de la variabilité de la température sur l'assimilation, le champ de température 2D a été lissé en utilisant différents filtres passe-bas.

Les résultats de chaque analyse ont été comparés aux états estimés par un filtre de Kalman classique (de dimension plus réduite, car opérant seulement sur la latitude et la longitude). Les résultats concernant le PF sont :

- Une meilleure estimation des positions à l'échelle locale grâce à l'assimilation du champ de température à sa meilleure résolution possible,
- Une correction satisfaisante des biais en latitude (constants ou variables), grâce à la structuration zonale du champ de température,
- Une interaction entre la précision de la mesure de géolocalisation et celle du champ de température utilisé dans l'assimilation, favorisant l'indétermination de la solution (probabilités multimodales couvrant plusieurs masses d'eau),
- Une difficulté du filtre à différencier les distributions de vitesses, et donc à estimer l'alternance entre modes de déplacement.

Cette étude permet d'identifier un écueil majeur de la valorisation des enregistrements de marques : il s'agit de définir exactement le schéma de biais/erreur affectant tous les processus d'observation. Ceci inclut la géolocalisation, mais aussi le champ de température employé dans l'ajustement. Or ce champ de température peut être issu 1) d'une cartographie de surface mesurée par radiométrie, 2) d'une interpolation de profils bathythermiques, ou 3) de modèles de circulation. Chaque source possède des limitations et des avantages précis : une interpolation par analyse objective des trois sources pourrait fournir une information de qualité suffisante pour l'analyse des données de marques.

Type	+	-
Radiométrie	Faible biais, couverture	Nuages, 2D
Profils T-Z interpolés	Faible biais, vertical	Faible couverture, 1D
Modèles de circulation	3D, couverture	Biais, résolution

Tableau 1. Données auxiliaires aux marques archives

3.5. Application à un cas réel

L'utilisation conjointe d'enregistrements d'archives et des sorties d'un modèle de circulation est ici illustrée dans la région du Gulf Stream : cette région présentant une forte variabilité océanographique, on s'attend à observer cette variabilité dans les enregistrements de marques (en termes de mesures de température, voire en terme de comportements individuels sur la verticale). Toutefois la grande dispersion des géolocalisations, le nombre important de données manquantes et le schéma de biais/erreur complexe du modèle dans cette région font de ce scénario un cas difficile.

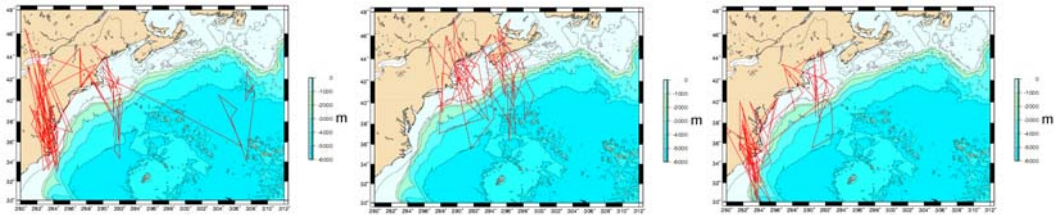


Figure 3-7. Géolocalisations estimées par analyse des photomésures des marques 37008, 37011 et 37006

Les enregistrements de trois marques de type « pop-up », fournis par Dr Lutcavage (University of New Hampshire) ont été considérés pour cette application. Ils sont issus d'une campagne de marquage dans le Golfe du Maine (juillet 2002 – janvier 2003) au cours de laquelle 67 individus de 100 à 300 kg ont été équipés. Un net mouvement vers les côtes de Virginie et de Caroline du Nord a été observé à l'automne (Wilson et al. 2004), sans toutefois toucher les zones de reproduction connue du détroit de Floride et du Golfe du Mexique. Un comportement vertical complexe, corrélé en particulier avec la présence de filaments et de tourbillons chauds du Gulf Stream ou la stratification des eaux côtières a été mis en évidence. Le grand nombre de données globalement disponibles permet de définir de bonnes statistiques de couverture spatiale dans la zone d'étude : toutefois la valorisation **individuelle** des marques est ici problématique. De grandes plages de données sont manquantes, tant sur les enregistrements à haute fréquence (1 heure) de profondeur et de température, que sur les géolocalisations journalières (liées à une indétermination de la durée du jour). La figure 3-8 montre le taux de remplissage des séries de mesures pour chaque marque.

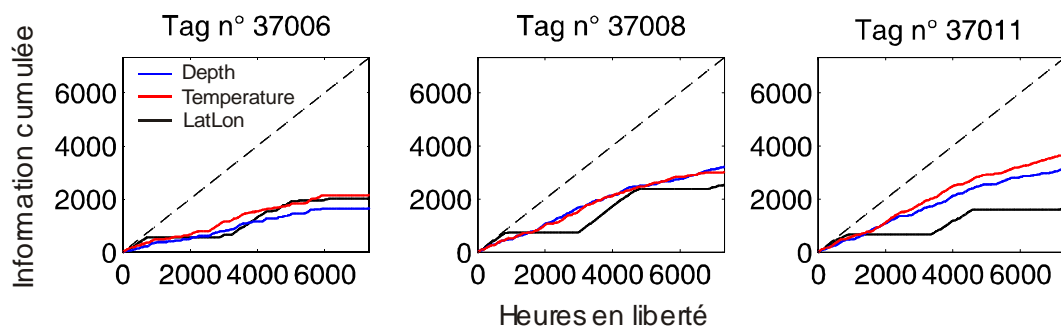


Figure 3-8. Informations cumulées (en points par heure) des trois types de mesure

Il apparaît toutefois que les observations de température et de profondeur sont les plus régulières. Le signal température-profondeur reste particulièrement variable pour les marques 37006 et 37008, comme le montre leurs diagrammes T-Z respectifs.

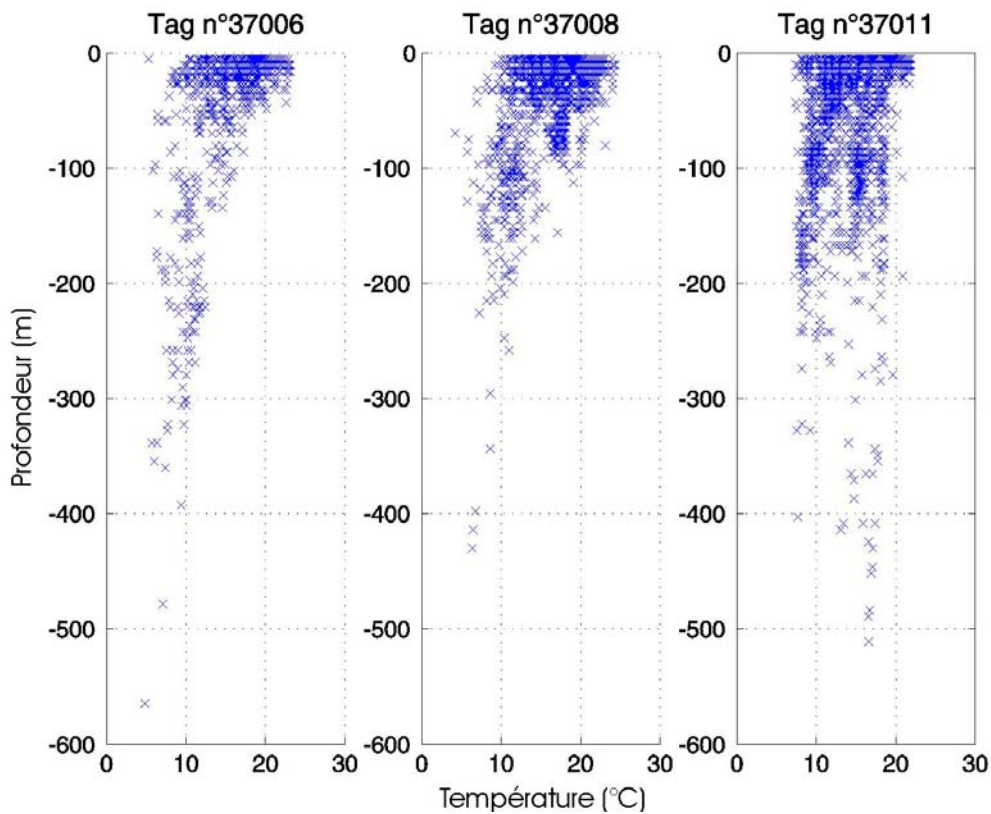


Figure 3-9. Profils T-Z des marques 37006/37008/37011

Une première analyse basée sur le filtre de Kalman simple (Sibert and Nielsen 2004) a été menée, autorisant l'estimation des paramètres de diffusion et de biais, tels que proposés par Sibert et al. (2003). Les itérations effectuées sur la marque 37008 n'ayant pas convergé vers une solution réaliste, son ajustement a été effectué avec les paramètres de la marque 37011.

Marque	U	V	D	Bx	By	Sx	Sy	A0	B0
37006	-0.3	-0.44	217.4	-0.14	0.07	1.13	2.01	0.24	20
37008	-	-	-	-	-	-	-	-	-
37011	0.28	-0.25	251.3	-0.7	0.09	1.42	2.8	0.69	23.2

Tableau 2. Paramètres d'advection et de diffusion estimés par filtrage de Kalman, d'après l'implémentation KFTRACK (Sibert and Nielsen 2004)

Les paramètres concernant les mouvements advectifs et diffusifs sont très comparables pour les deux trajectoires ajustées (37006 et 37011), qui semblent rendre compte d'un comportement relativement stationnaire, ou d'un mouvement en boucle. On s'attend, dans des cas plus hétérogènes, à ce qu'une grande partie de la variance ajustée soit attribuée au comportement diffusif, le terme constant d'advection ne pouvant capturer des mouvements très complexes comme le cas n°37011. Ceci pourrait expliquer en partie la non-convergence du schéma d'estimation.

Malgré sa robustesse, ce filtre ne permet pas de rejeter des positions aberrantes (e.g. sur terre), ou de bénéficier de l'information fournie par les mesures de profondeur (la bathymétrie devenant contraignante si l'on connaît le comportement vertical à la côte du poisson), ou de température.

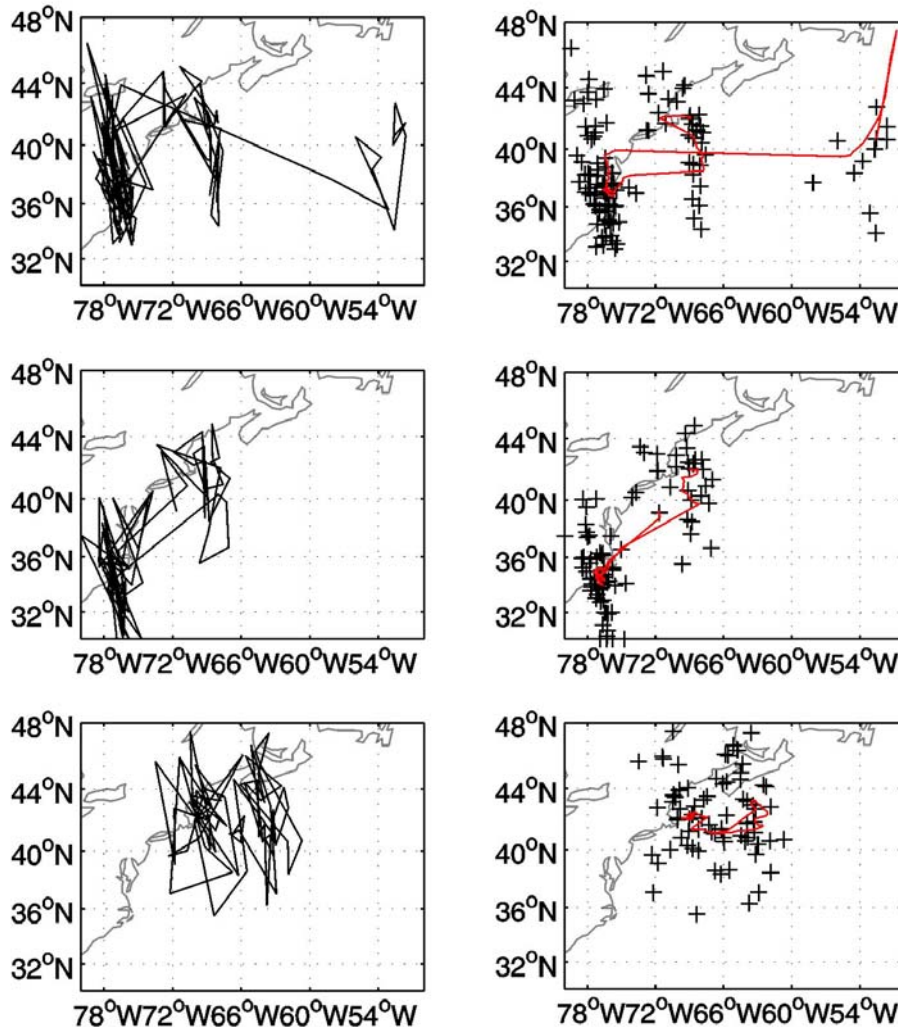


Figure 3-10. Trajectoires brutes (à gauche) et réestimées par filtrage de Kalman (Sibert et al. 2003) des marques 37008, 37006 et 37011

Le filtre particulière a donc été employé ici pour tenter d'extraire un supplément d'information de ces données. Le modèle Mercator sous sa configuration PSY2v1R1 (code OPA assimilé en altimétrie) a été utilisé. Avec l'aide du bureau Utilisateurs du GMM Mercator, 69 fichiers agrégés à 5 jours ont été extraits des produits vitrines au 1/8^{ème} de degré, couvrant la zone 80W/40W/30N/40N et la période d'août 2002 à juin 2003. Les champs moyens de température et salinité sur cette période ont été extraits, ainsi qu'un exemple du flux u-v de surface à la date du 1^{er} août 2002.

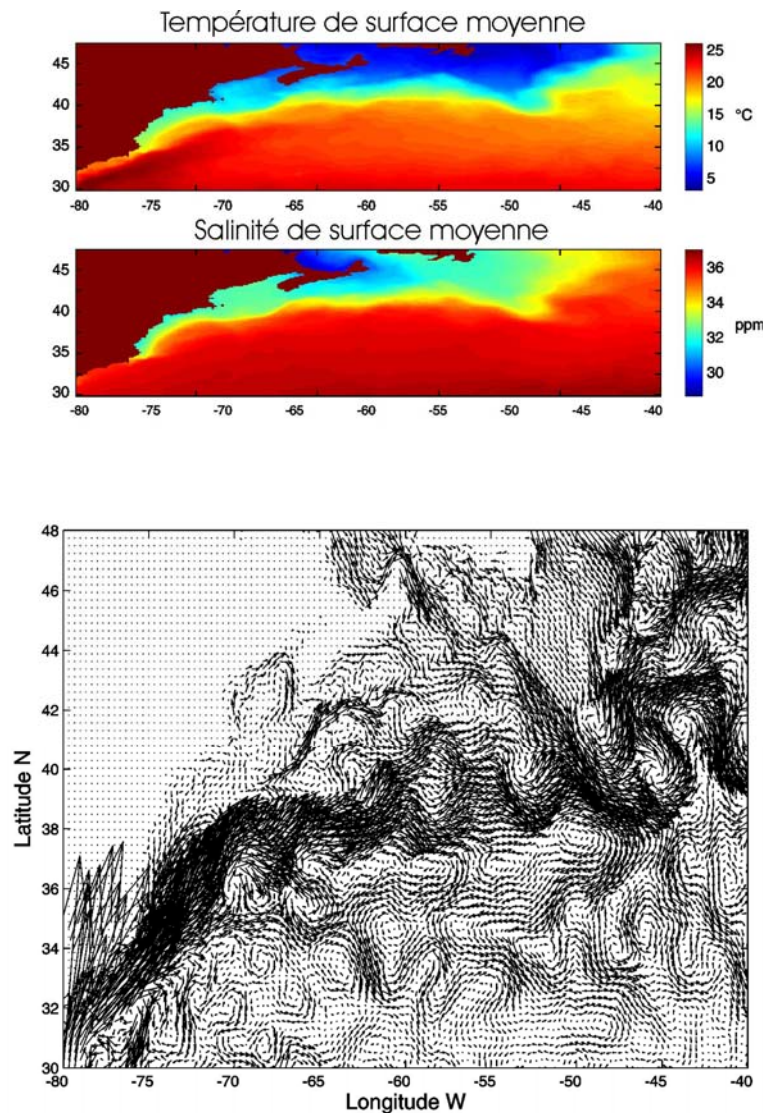


Figure 3-11. Température et salinité de surface moyennes PSY2v1 sur la période août 2002-juin 2003, et exemple du flux de surface au 1er août 2002.

Le schéma de biais/erreur du modèle fait l'objet d'un suivi constant (Etienne and Charon 2004, Greiner et al. 2004): des comparaisons prédiction/observation au niveau de mouillages ou de radiales précises permet de cartographier les écarts aux observations. La comparaison de moyennes annuelles avec la température globale interpolée de Reynolds (Reynolds and Smith 1994) montre un biais chaud important (jusqu'à 2 degrés) à la limite nord du Gulf Stream, et des biais froids généralisés au sud de son enveloppe.

Les profils bathythermiques au niveau du Gulf Stream révèlent que ce biais chaud est maximum au-dessus de 100m (-2°C à 75m), et minimum entre 150 et 300m. Ces biais ne sont pas stationnaires, et dépassent 3°C en hiver pour cette version de PSY2v1 (Etienne and Charon 2004). L'isotherme 7°C est ainsi bien trop profonde dans le modèle (600 à 700m contre 400m dans les observations). Les couches supérieures ne seraient donc en général pas assez stratifiées.

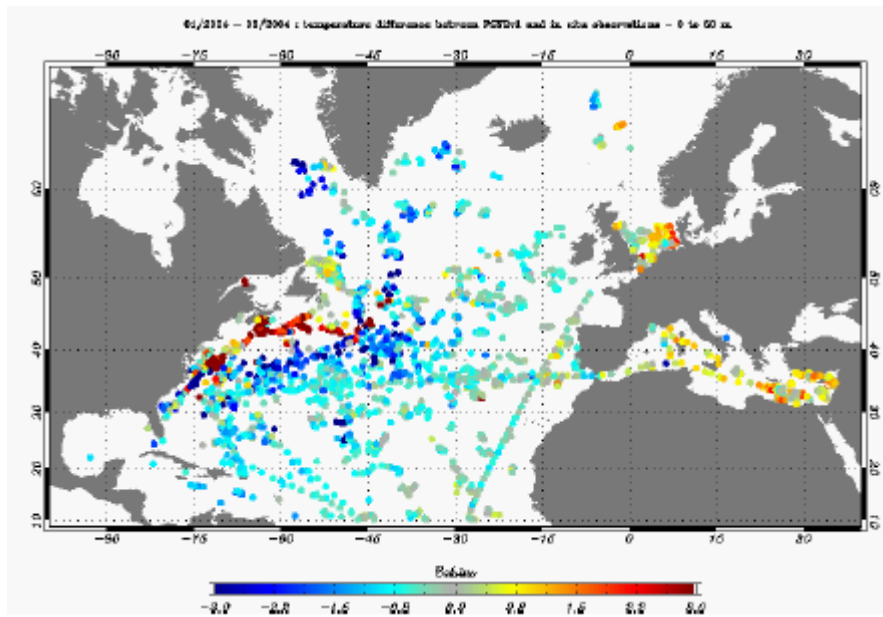


Figure 3-12. Biais de température entre PSY2v1 et observations *in situ* dans la zone -0 à 50m

L'inclusion du champ de température dans notre filtre Particulare, si elle fournit une information supplémentaire importante, doit donc être fait avec précaution: la grande variabilité verticale observée dans les marques (générant des variations de plus de 10°C parfois) est sans aucune commune mesure avec la variance spatiale de la température de surface. La cartographie des biais PSY2v1/Reynolds n'étant pas disponible pour cette étude, la variabilité de la température *in situ* (mesurée par la marque et non résolue par le modèle), l'erreur du modèle et son biais systématique ont été modélisés par une erreur unique sur la température. En première approximation, des erreurs gaussiennes sur la température ont été spécifiées, avec un écart-type de 0.5 à 1°C (autorisant ainsi la possibilité d'écarts importants). Les erreurs de localisation ont été fixées à partir des itérations de Kalman précédentes (*i.e.* 1 et 2 degrés en longitude et latitude, respectivement).

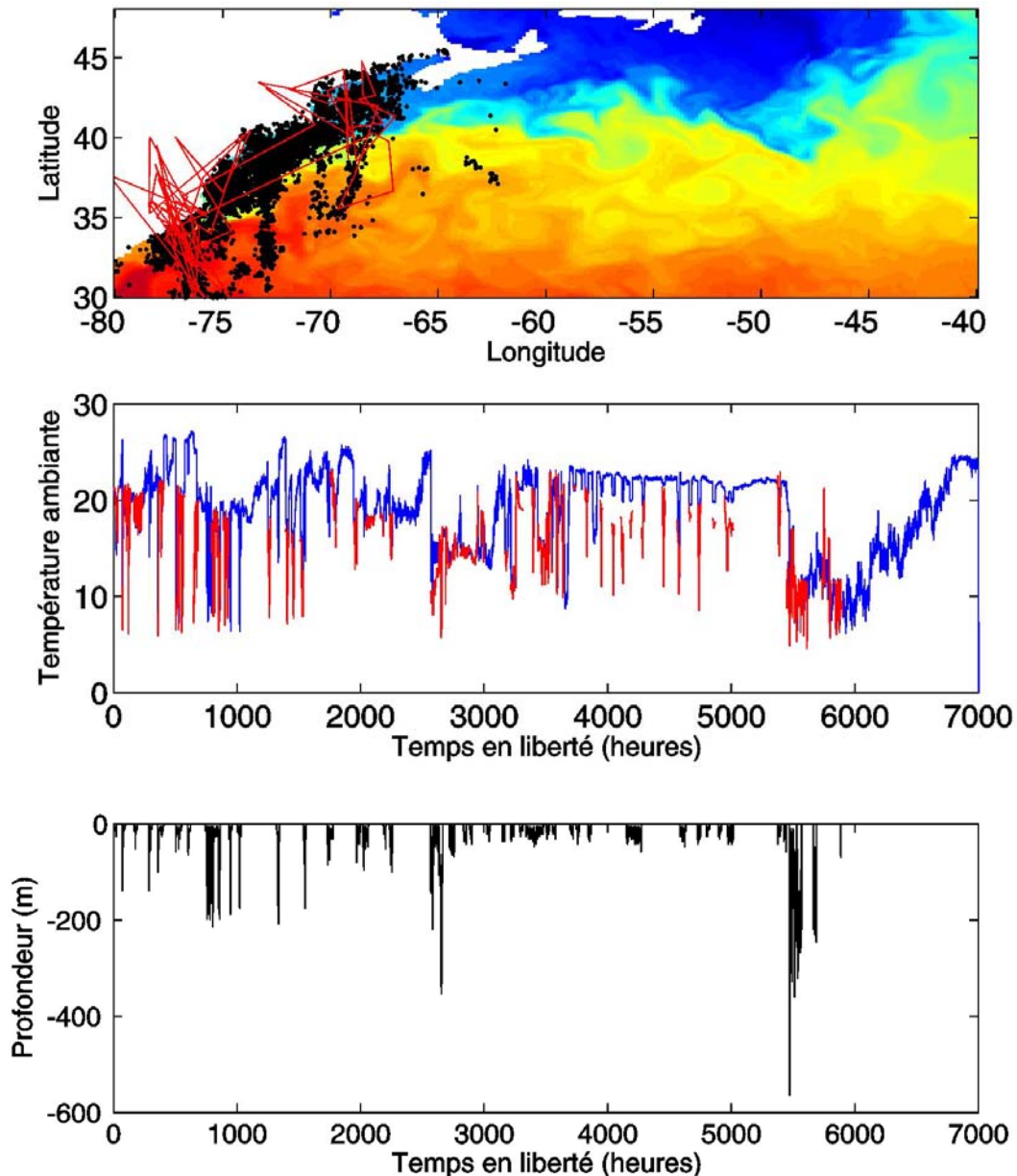


Figure 3-13. Haut : exemple de divergence d'états estimés par un filtre particulaire sur la marque 37006 (en noir : particules, en rouge : observations). Milieu : dans l'espace température (en rouge: observation, en bleu: estimation). Bas : profondeur observée.

La plupart des ajustements ont malheureusement été insatisfaisants, avec une extrapolation souvent irréaliste de la géolocalisation des marques. Si la température est globalement contraignante (*i.e.* la plupart des événements de plongée sont par exemple bien reconstitués au début), il apparaît que le filtre particulaire diverge constamment en longitude et latitude. Des variations importantes dans l'estimation des données manquantes sont ainsi visibles, et ne correspondent pas *a priori* à des états réalistes. La divergence du filtre est en particulier visible de H:3800 à H:5000 où la série de plongées correspond visiblement à des états non estimables (l'amplitude thermique y atteint 10°C, mais les zones proposées par le filtre n'y autorise que des amplitudes minimales de 2 à 3°C).

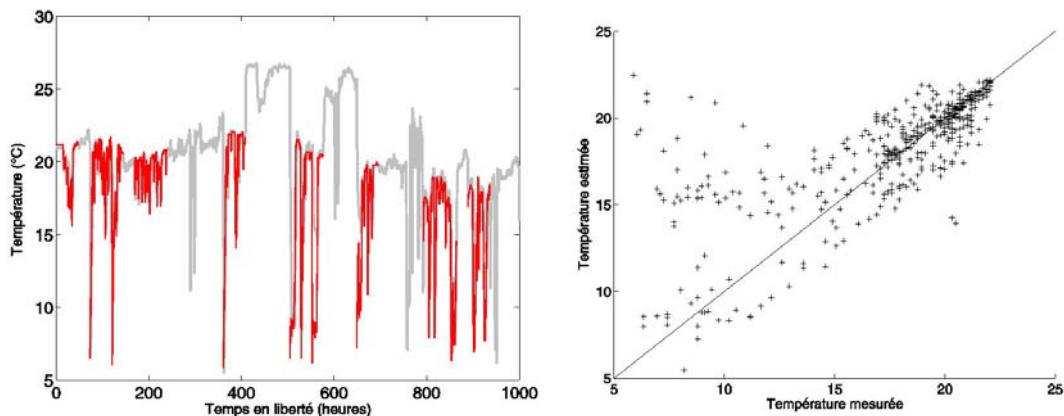


Figure 3-14. Zoom sur une portion du profil de température estimée, et graphe de dispersion mesure/estimation dans l'exemple de la marque 37006

Le filtre Particulaire proposé dans ce travail, s'il semble performant lorsque le problème est "bien posé", manque en fait de robustesse: son comportement Markovien, par définition, lui accorde une souplesse trop grande dans l'estimation des états à $t+1$ et facilite sa divergence. L'information disponible pour son ajustement étant incomplète, bruitée, biaisée et spatialisée, il a été impossible de réduire les erreurs convenablement en utilisant la version actuelle du filtre. Les matrices de covariance d'erreur en température sont en effet complexes, et distribuées dans les quatre dimensions (x,y,z,t) . Le filtre Particulaire, ou Markov Chain Monte Carlo (MCMC) séquentiel, a été ici implémenté dans une version simple, *i.e.* la densité de probabilité du *prior* est simplement la densité estimée au pas de temps précédent. Intuitivement, la trop grande souplesse du filtre peut être attribuée à son caractère Markovien, *i.e.* un "oubli" trop rapide des dynamiques passées pour ne garder que les états et paramètres les plus proches dans le temps. En d'autres termes, la constante de temps de la dynamique implémentée est beaucoup trop faible: le filtre n'emploie pas (ne valorise pas) l'information de l'ensemble de la série temporelle, mais travaille à court terme seulement. Il est en cela fondamentalement différent d'un interpolateur (une remarque: en effectuant les itérations dans le sens rétrograde, il serait possible d'améliorer l'estimation par lissage, *i.e.* par l'emploi des observations passées ET futures). Ce problème a déjà été identifié pour les filtres MCMC ou PF, et les solutions proposées sont toutes assez similaires: il s'agit d'améliorer l'estimation en utilisant un meilleur *prior*, *e.g.* une version lissée de la série d'observation, voire une série d'états estimés par un filtre de Kalman (travaillant sur la version linéaire du système). Ces solutions ont été très récemment étudiées et détaillées par Dahia (2005), qui propose en particulier la solution du filtre de Kalman-particulaire.

3.6. Perspectives immédiates

Le succès du traitement des enregistrements d'archives doit être relativisé: leur validation à l'aide de capteurs tiers (e.g. GPS) reste problématique pour des animaux à branchies (Teo et al. 2004). Si les méthodes proposées sont efficaces dans des cas contrôlés ou pour des situations relativement simples (Nielsen 2004), les erreurs d'observation (généralement corrélées) des données de terrain les mettent largement en défaut.

Un problème de résolution? Il apparaît que l'amélioration des *a priori* sur les erreurs de cartographie des champs de température soit nécessaire, en plus de l'amélioration de la méthode d'estimation. L'assimilation des mesures et profils de température dans la version PSY2v2 laisse espérer de meilleures performances: une réduction des biais de température en surface ET une meilleure représentation de la thermocline sont nécessaires à la valorisation des enregistrements de marques archives. Ceci est particulièrement vrai pour des espèces dotées de thermorégulation comme le thon rouge ou le thon obèse, *Thunnus obesus* (l'essentiel de la variabilité en température étant due à la dimension verticale et générant des erreurs spatialisées). L'association OGCM-prédateurs pélagiques reste une piste prometteuse, et des problèmes durs (e.g. de couplage eulérien/lagrangien) restent à résoudre.

Le problème technologique. Malgré tout, l'étude du comportement d'espèces pélagiques par marques archives reste non-triviale: l'interaction entre variabilité de la dynamique individuelle, variabilité environnementale et erreurs de mesure mène à des scénarios de modélisation difficiles. Il semble raisonnable de s'orienter vers le développement de nouveaux capteurs dédiés à la géolocalisation seule. De nouveaux modes de navigation pourraient être envisagés, comme la navigation à l'estime: la mesure de l'azimut de la marque par compas magnétique à trois axes, combinée à la mesure de la vitesse de nage (un simple tube de pitot pourrait faire l'affaire), ou à l'intégration de mesures d'accélération grâce à un accéléromètre piézo-électrique. Des technologies peu gourmandes en ressources restent primordiales (l'aspect stockage mémoire devenant de moins en moins critique aujourd'hui). La solution de l'accéléromètre, si elle peut générer des dérives, peut fournir des informations de grande valeur sur les stratégies de nage du thon rouge (e.g. "burst and glide"), et la fréquence d'évènements comme la chasse ou la fuite vis-à-vis de prédateurs (Videler and Weihs 1982, Newlands et al. 2004).

*La mise au point de nouveaux types de capteurs autorisant la navigation à l'estime doit être encouragée. Le **manuscrit 4** propose quelques voies de recherche, basées e.g. sur l'emploi de compas magnétiques, d'accéléromètres de filtres polarisants.*

De la question initiale. De nouveaux capteurs pour les marques "pop-up" pourraient mener à de nouveaux types de travaux en éco-éthologie: si les marques archives proposer de mesurer la température interne (et renseignent sur la fréquence d'ingestion de proies ou sa dynamique de thermorégulation) leur implantation chirurgicale reste contraignante, et nécessite de recapter le poisson. Traditionnellement, on tente de croiser l'histoire géographique du poisson (e.g. la fréquentation de certaines masses d'eau) à son histoire écologique (e.g. la possibilité qu'il vienne s'y reproduire ou s'y nourrir). Or le "biologging" est actuellement un champ de recherches largement orienté par la technologie, les problèmes d'implémentation de capteurs et la question de leur fiabilité (Sims 2003). Pour progresser à long terme dans ce champ de recherches, il semble important de privilégier la mise au point de marques desservant l'objet du scientifique, plutôt que de placer des outils génériques sur le marché. Les découvertes "évidentes" et spectaculaires (*i.e.* les travaux rapportant l'étendue surprenante de la migration de certains individus) ont un fort impact sur la communauté scientifique, mais elles perdent aujourd'hui peu à peu leur aspect de nouveauté (Block et al. 2001, Block et al. 2002), et doivent mener à des travaux plus orientés. Aborder l'exploration de l'habitat pélagique par les thonidés comme un problème inverse (*i.e.* caractériser sa structure sachant le comportement de l'individu) apparaît en particulier comme une voie prometteuse, bien que gourmande en données (Gutenkunst et al. 2004).

Changement d'échelle. Un aspect intéressant du traitement du signal a été abordé dans ce chapitre, *i.e.* la contribution des différentes échelles d'espace et de temps au processus d'estimation. En d'autres termes, l'information globalement présente (à grande échelle) dans un jeu de données peut servir à estimer localement des états: la variabilité locale (haute fréquence) est alors considérée comme peu informative (peu contraignante) et donc traitée comme un bruit d'observation. Ceci revient à interpoler le signal. Si le problème est numériquement trop complexe, on se tournera vers le filtrage ou le lissage (de préférence récursif): la simple mise à jour d'une matrice de covariance ou d'un ensemble de points est alors en général suffisante. Ceci est très analogue au problème d'estimation en dynamique des populations: le but est d'alors d'extraire l'information écologique présente dans une série temporelle, en décomposant sa variance en tendances, cycles et résidus. Nous changeons d'échelle pour la suite de ces travaux, et abordons les fluctuations de la population de *Thunnus thynnus*.

4. Nouveau regard sur les fluctuations de la population de *Thunnus thynnus*

Différentes méthodes d'estimation et de test d'hypothèse ont été proposées dans les travaux précédents: elles sont ici appliquées aux séries historiques de capture de thon rouge (Ravier and Fromentin 2001), afin d'approfondir l'analyse exploratoire menée dans la thèse de C. Ravier (Ravier 2003). Les hypothèses proposées par ces auteurs sont en particulier réexaminées et étayées, au regard de nouveaux résultats sur la nature des fluctuations de *Thunnus thynnus*. L'accent est résolument mis sur la recherche du signal écologique, la détection d'un déterminisme ou tout autre signature de mécanismes non aléatoires, et le détail de la dynamique des séries temporelles de capture (**manuscrit 5**). Un pattern complexe de densité-dépendance est mis en évidence, ainsi qu'un possible forçage par le climat à grande échelle. L'approche fréquentielle (*i.e.* le domaine de Fourier) est en particulier détaillée au cours du chapitre et illustrée dans un cas simple (**manuscrit 6**).

4.1. La base de données madragues

Sur le pourtour Méditerranéen et à Gibraltar, une pêcherie traditionnelle cible le thon rouge et tire profit de son comportement de migrateur: les **madragues**, engins de pêche ancrés au rivage, interceptent les poissons puis les concentrent dans une chambre où a lieu la capture. Le coût d'entretien de ces structures, la main d'œuvre nécessaire à leur opération et au conditionnement des captures ont réservé la propriété des madragues à la noblesse locale, aux notables et aux marchands méditerranéens (Farrugio 1981, Doumenge 1998). Les livres de comptes des administrations et les archives familiales renferment ainsi une

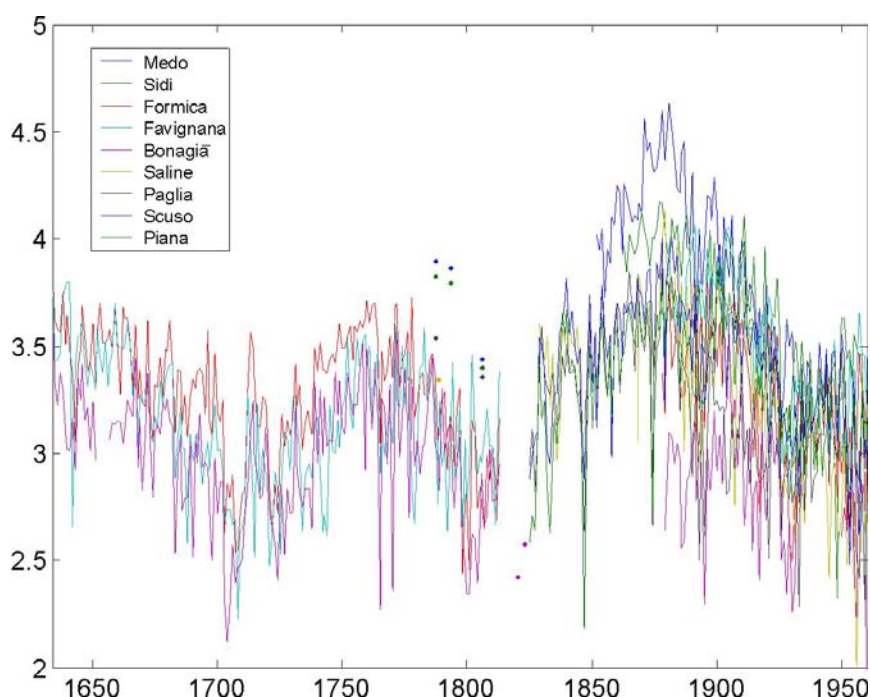


Figure 4-1. Les neuf séries temporelles utilisées pour cette étude

information précieuse sur la population du thon rouge et ses captures historiques. De récents efforts de collecte et de synthèse bibliographique ont permis la construction d'une base de données de capture importante couvrant plus de trois siècles (Ravier and Fromentin 2001, Fromentin 2002). La faible évolution technologique de l'engin et l'étude des synchronies entre séries de capture sur le pourtour du bassin (Fromentin et al. 2000) ont montré comment ces séries pouvaient refléter les variations d'abondance de la population de *Thunnus thynnus* (*i.e.* des fluctuations de capturabilité – liées au mode de pêche, et donc de moindre intérêt écologique – sont alors peu probables) (Ravier and Fromentin 2001).

Les captures montrent un spectre de variabilité assez clair, dont 50% de la variance est expliqué par des pseudo-cycles de périodes 100-120 ans et secondairement de 15-30 ans (Ravier and Fromentin 2001). Une hypothèse, dérivée de considérations théoriques sur les traits d'histoire de vie du thon rouge, a été avancée pour expliquer ces cycles à moyen terme (~20 ans). En particulier, l'interaction d'un grand nombre de classes d'âge au moment de la reproduction (de l'ordre de 15) est susceptible de générer des signaux distribués sur de nombreuses cohortes, et des variations d'abondance durables (Fromentin and Fonteneau 2001, Bjornstad et al. 2004). Une hypothèse climatique (*i.e.* impliquant l'environnement sur de grandes échelles de temps) a été proposée pour expliquer les cycles de grande longueur d'onde (~100-120 ans): (Ravier and Fromentin 2004) ont mis en évidence de fortes relations négatives et systématiques entre les captures de madragues et différents séries de température de l'hémisphère nord (instrumentales ou reconstruites). Ce résultat, valide sur la période 1850-1950, a conduit ces auteurs à postuler un effet du climat océanique sur les migrations des reproducteurs. L'hypothèse alternative, celle d'un effet sur la phase recrutée (Santiago 1998, Marsac 1999), est *a priori* moins vraisemblable car une relation positive serait alors attendue. La mise en relation avec d'autres indices environnementaux régionaux comme la NAO ou la LOD n'a pas indiqué de lien apparent (Ravier and Fromentin 2004).

L'hypothèse actuelle repose donc sur un effet du climat sur la dynamique migratoire du thon rouge Atlantique (et donc son accès aux zones connues de reproduction). La difficulté d'obtention de séries environnementales à long terme rend le test et l'approfondissement malaisés. Les phénomènes incriminés opérant à basse fréquence (donc générant un signal fortement autocorrélé), la plupart des méthodes statistiques classiques sont mises en défaut lors du test de significativité (Student 1914, Bartlett 1946, Fortin and Jacquez 2000). En particulier, il semble difficile d'isoler les mécanismes précis à l'origine de ces fluctuations et de proposer une hypothèse nulle précise (*e.g.* variations de la circulation océanique de surface, variations de production biologique régionales...). Le test d'hypothèse pour une dynamique basse fréquence devient ainsi doublement problématique: on dispose rarement de plusieurs réalisations du même processus, ce qui augmente le risque de

corrélations accidentelles par e.g. la mise en relation de tendances fortuites (Ravier and Fromentin 2004).

L'objectif est ici de préciser la **structure temporelle** des fluctuations de la population au moyen de méthodes exploratoires particulières (e.g. issues du traitement de signaux chaotiques), puis de construire un modèle de leur dynamique. Le problème du **test d'hypothèse** est ensuite abordé, en basant la réflexion sur une version réduite de la base de données originelle. Neuf séries parmi les plus longues (>80 ans) sont retenues: elles sont issues des sites siciliens (madragues Favignana, Formica et Bonagia), de Sardaigne (Saline, Porto Paglia, Porto Scuso, Isola Piana), de Tunisie (Sidi Daoud) et du Portugal (Medo das Casas).

Nom	Localisation	Période	Durée	Médiane	%CV	β^3
<i>Medo das Casas</i>	Portugal	1852-1933	112	8.90	10.57	1.22
<i>Sidi Daoud</i>	Tunisie	1863-1960	98	8.36	10.09	1.10
<i>Formica</i>	Sicile	1634-1813 1878-1960	302	7.58	9.02	0.85
<i>Favignana</i>	Sicile	1634-1813 1878-1960	302	7.56	10.74	1.07
<i>Bonagia</i>	Sicile	1657-1809	249	6.82	10.73	0.64
<i>Saline</i>	Sardaigne	1868-1960	114	7.50	12.19	0.91
<i>Porto Paglia</i>	Sardaigne	1825-1960	138	7.37	10.31	0.64
<i>Porto Scuso</i>	Sardaigne	1825-1960	139	7.95	8.85	0.69
<i>Isola Piana</i>	Sardaigne	1825-1960	139	7.73	9.16	0.87

4.2. Methodes non-lineaires

Construction d'un espace des phases. Le modèle sous-jacent à la dynamique d'une population peut s'écrire sous sa forme générale:

$$X(t+1)=f[X(t),U(t),w(t)],$$

où U est une variable forçante et w la composante stochastique du système. En pratique, f est généralement inconnue, et les variables d'état X ne sont mesurées qu'en partie (e.g. $Y=g(X)$ et $\dim(Y)<\dim(X)$). Une suite de mesures Y est donc seulement disponible. Le modèle pourrait alors être construit sous la forme suivante (Turchin 2003):

$$Y(t+\sigma_p)=F[Y(t),Y(t-\sigma), \dots, Y(t-(d-1)\cdot\sigma), U^1(t),U^2(t), \dots, \varepsilon(t)]$$

On peut noter que le processus d'observation (l'échantillonnage) est partiellement spécifié sous la forme de σ , qui constitue le **délai d'opération** du modèle, i.e. l'écart nécessaire entre deux mesures pour extraire un maximum d'information sur X (au sens algorithmique de Shannon). Les paramètres σ_p et d constituent l'horizon de prédiction et

³ La valeur β synthétise le ratio haute/basse fréquence d'une série temporelle, lorsque son spectre de variance est approché par une relation $1/f^\beta$. Les valeurs calculées ici concernent les séries non détendancialisées.

l'ordre (la dimension) du processus, respectivement. Les erreurs de processus (liée à l'intervention possible d'autres variables ou à une structure fondamentalement stochastique) sont synthétisées par ε . Enfin, on définit par U^i une série de covariables (e.g. forçage par l'environnement, « contrôle » par la pêche) afin de compléter l'analogie avec la notation des systèmes dynamiques. La forme de F étant inconnue, nous sommes dans le domaine de l'**exploratoire**. En écrivant le modèle ainsi, on suppose qu'il est possible de **reconstruire** cette dynamique F en s'intéressant uniquement à l'évolution de l'observation Y . Ceci peut se concevoir intuitivement: SI X est effectivement soumis à une dynamique (i.e. c'est un signal non-aléatoire), ALORS la mesure d'états successifs Y_t contient une information sur cette dynamique. Il nous appartient alors, muni des outils de l'écologie quantitative, de l'extraire et de l'interpréter. Une remarque intéressante est que ce raisonnement (à temps discret) peut être étendu à sa version continue: il suffit d'exprimer le signal X en fonction d'une suite de ses dérivées partielles (celles-ci pouvant être calculées empiriquement par lissages successifs). En pratique, cette méthode s'apparente à la méthode des **moments**, souvent utilisée dans sa version de deuxième ou troisième ordre pour la modélisation de trajectoires lagrangiennes (Castellari et al. 2001, Özgökmen et al. 2001, Piterbarg and Özgökmen 2002). Son application reste toutefois difficile lorsque le rapport signal-bruit est trop élevé, le calcul d'une dérivée étant extrêmement sensible aux bruits d'observation (Özgökmen et al. 2000).

Pour la suite du raisonnement, la notation peut être simplifiée ainsi: $X(t)=F(X(t-\sigma), X(t-2\sigma), \dots, X(t-d\sigma), \varepsilon(t))$. Les covariables U influant sur la dynamique d'une population étant souvent mal connues et/ou mal mesurées, leur étude sera abordée en fin de chapitre. La question se résume alors à extraire la forme **déterministe** de F (aux erreurs ε près) en ne se basant que sur une suite d'observations. Les travaux de **Takens** (1981) stipulent que ceci est possible sous certaines hypothèses.

Théorème de Takens

Soit un système dynamique f opérant sur un attracteur A tel que:

- A soit de dimension d_A
- A contient un nombre fini de points d'équilibre,
- A ne contient pas d'orbites périodiques de période σ ,
- A contient un nombre fini d'orbites de période $p\sigma$, avec $3 < p < m$,

Alors quelle que soit la fonction de mesure $h(x)$, et pour $m \geq 2d_A + 1$,

$F = h(x_t), h(x_{t-\sigma}), \dots, h(x_{t-(m-1)\sigma})$ définit une **reconstruction** de la dynamique f , et possède les mêmes caractéristiques géométriques (e.g. dimension de l'attracteur A , exposants de Lyapunov...)

Ce théorème est valide pour les dynamiques **déterministes** (ou au moins pour $\varepsilon(t) \ll |X(t) - X(t + \Delta t)|$, donc des faibles valeurs de bruit), et stipule qu'il n'est pas nécessaire d'écrire une dynamique sous sa forme réelle et complète pour l'étudier, mais qu'il suffit pour cela d'organiser les observations dans un espace des phases empirique, basé sur des **délais** précis. Le théorème de Takens ne nous donne malheureusement pas la

valeur des paramètres m et σ nécessaires à la reconstruction (Casdagli et al. 1991). En pratique, la valeur de m (la dimension de cet espace) doit être **suffisamment** élevée pour permettre une bonne reconstruction de l'attracteur originel (Grassberger et al. 1991). Le délai σ doit être suffisant pour éviter une autocorrélation trop élevée entre mesures successives, tout en restant suffisamment informatif sur la dynamique de f . Plusieurs méthodes empiriques peuvent aider à déterminer m et σ (Rosenstein et al. 1994, Solé and Bascompte 1995): Le calcul et le seuillage de l'autocorrélation peut ainsi servir à déterminer σ , alors que l'étude dite "des faux voisins" permet de déterminer la valeur de m (Kantz and Schreiber 1997). Il s'agit alors de reconstruire la trajectoire $X(t)$ dans un espace de dimension croissante: un certain nombre de points définis par $[X(t-\sigma), X(t-2\sigma), \dots, X(t-d\sigma)]$ seront potentiellement proches (au sens euclidien). Dans un système déterministe, deux points proches au temps t le resteront à $t+\Delta t$ (l'écartement variant proportionnellement avec le temps pour une dynamique linéaire). En augmentant m , on déplie pas à pas la trajectoire: lorsque ce processus ne permet plus de séparer des points proches, alors on peut affirmer que la dimension m permet de capturer la dynamique du système.

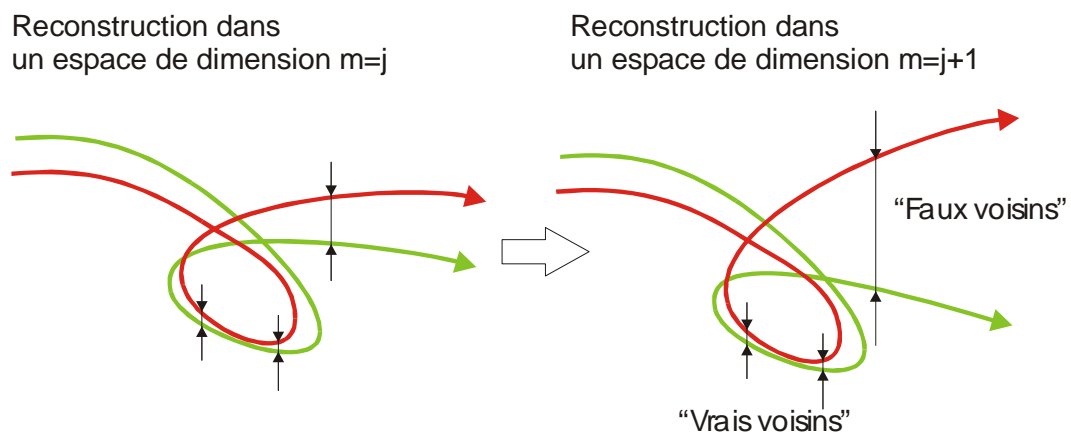


Figure 4-2. Illustration de la méthode des faux voisins et du changement de conformation de la trajectoire avec la dimension de Takens

En pratique, les systèmes écologiques étant souvent non-stationnaires, il peut être intéressant de surestimer la dimension de reconstruction m (Hegger et al. 2000). En effet la dynamique du système peut être de variance maximale dans un certain référentiel (*i.e.* un certain système d'axes orthogonaux) puis, lors d'un changement de régime, se voir reportée sur un autre référentiel, donc un nouvel ensemble de dimensions.

Cette courte introduction à la construction empirique de l'espace des phases ouvre la porte à de nombreuses possibilités. On peut en particulier étudier la **forme** de f (*i.e.* linéaire ou non-linéaire?) en appliquant la définition littérale de la linéarité. Il suffit alors d'étudier l'évolution de la distance $|F(X(t))-F(X(t+\Delta t))|$ pour conclure (Patil et al. 2000). Les exposants de Lyapunov peuvent par exemple être employés (Casdagli 1989).

Les exposants de Lyapunov

Lorsqu'un système voit son état initial perturbé, le taux d'accroissement de cette perturbation au cours du temps peut être caractérisée. La linéarisation de l'orbite du système est alors nécessaire au calcul. On écrit: au point X^* :

$$X(t) = X^*(t) + U(t),$$

Où $U(t)$ est la déviation moyenne par rapport à la trajectoire de référence (non perturbée). Les exposants caractéristiques de Lyapunov sont donnés par le théorème d'Osedelec:

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t} \ln |U(t)|$$

U se calcule à partir des Jacobiens le long de l'orbite (voir le chapitre 3). Il existe autant d'exposants de Lyapunov que de dimensions du problème: chacun caractérise alors le degré de divergence de la trajectoire du système le long de cette dimension (Eckmann and Ruelle 1985). Si le système est caractérisé par un équilibre stable, le plus grand exposant de Lyapunov sera négatif: une perturbation à l'équilibre s'atténuera nécessairement. Un cycle limite impliquera un exposant égal à zéro. Un attracteur chaotique présentera au moins un exposant positif.

Le degré de stochasticité du système peut aussi être étudié, en se reposant sur une définition très intuitive du **hasard**, qui est "l'absence ou la rareté de phénomènes récurrents" (Chaitin 2001). La probabilité qu'un segment de trajectoire soit proche d'un autre segment renseigne donc directement sur le déterminisme du système. Pour un système complètement stochastique, la trajectoire occupera l'espace entier au bout d'un temps fini (un bruit étant un processus de dimension infinie). En d'autres termes, la formulation basée sur les **délais** permet non seulement l'écriture quantitative de modèles autorégressifs, classiques en écologie quantitative (Box and Jenkins 1976), mais aussi l'analyse qualitative de la dynamique cachée. Ce formalisme apparaît très adapté à l'écologie marine, qui fournit des systèmes presque toujours incomplètement mesurés. Notre cas d'application est aussi un système partiellement observé, puisque seules des mesures d'abondance sont disponibles (non détaillées en classes d'âge ou de taille) .

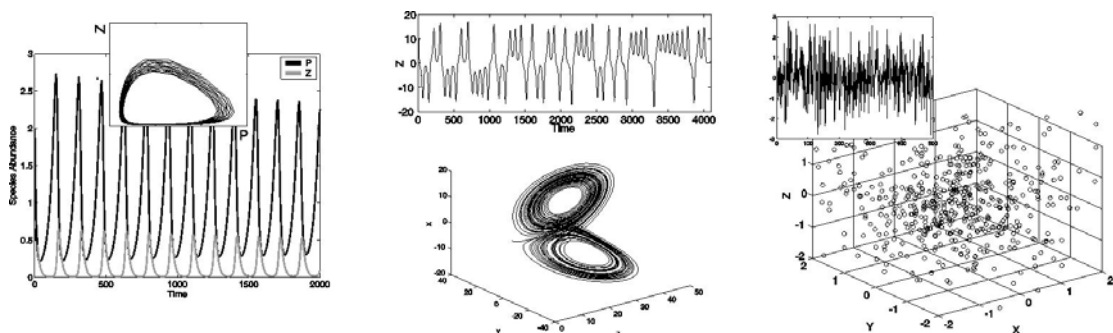


Figure 4-3. Illustration de l'espace des phases original de trois modèles a) déterministe (un modèle NPZ – voir article 4), b) chaotique (le système de Lorenz), c) stochastique (un bruit blanc).

Deux problèmes se profilent alors: 1) il est nécessaire de définir les statistiques adéquates et le protocole de leur test, afin de comprendre les mécanismes de la dynamique cachée f, 2) si la présence de bruits de

processus n'est *a priori* pas contraignante (elle renforce au contraire le caractère stochastique de f , et peut donc être facilement testée), *quid* d'un bruit d'observation?

Erreurs

L'approche déterministe d'un système suppose une formulation sans erreur du modèle sous-jacent, et permet la répétabilité des calculs et la prédiction de tout état $X_{t_0+\Delta t}$ sachant les conditions initiales X_{t_0} , pour un intervalle de temps Δt arbitraire (hors du domaine chaotique). L'approche stochastique, si elle repose malgré tout sur un noyau déterministe (Gurney and Nisbet 1998), autorise la prise en compte d'**erreurs** de nombreux types.

On distinguera les **erreurs de mesure**, généralement additives et ne modifiant pas la dynamique du système, des **erreurs d'estimation**, liées au traitement subséquent de la donnée pour en extraire l'information. Les **erreurs de processus** peuvent être liées à un manque de connaissance objective du système, et modélisée, faute d'un mécanisme candidat, comme une variable aléatoire de loi appropriée. Elles peuvent être aussi inhérentes au système, certains processus microscopiques fondamentalement stochastiques ne pouvant être réduits (*e.g.* les interactions moléculaires liées à l'agitation thermique). Ainsi la stochasticité démographique peut produire, dans un environnement stable, une loi de type binomial négatif (Kendall 1949), alors qu'on attend de la stochasticité environnementale une loi de variance proportionnelle au carré de la moyenne (Engen and Lande 1996).

Enfin, il convient de rappeler que, malgré les grandes avancées technologiques en calcul scientifique, nous sommes à ce jour dépourvu des capacités calculatoires analogiques et quantiques du réel... Ceci donne une valeur ambiguë aux **erreurs d'implémentation**, dont la minimisation est souvent vue comme le reflet d'un processus de réflexion rigoureux, malgré une contribution secondaire à la réflexion scientifique elle-même.

Du bruit d'observation. Avant de procéder, une courte description sur les effets du bruit dans l'espace des phases est nécessaire. On rappelle que la notion de **bruit** désigne toute dynamique stochastique non réductible à un système restreint d'équations. Un bruit d'observation (*i.e.* n'affectant pas la dynamique sous-jacente du système) aura pour effet: 1) une destruction de l'auto-similarité de l'attracteur si celui-ci est chaotique, 2) une dimension apparemment élevée de l'espace des phases aux faibles échelles (rappelons qu'un bruit est de dimension infinie), 3) une divergence des trajectoires voisines de type diffusif et non exponentiel (Kantz and Schreiber 1997)... Toute tentative d'estimation de f sera nécessairement affectée par les erreurs d'observation (Sivakumar 2000). Les effets du bruit de mesure ne peuvent donc être ignorés dans la modélisation d'un système caché. En pratique, il conviendrait d'estimer ses caractéristiques *a priori*, afin de s'affranchir d'interactions possibles entre les différentes étapes d'estimation (on remarquera que le cadre Bayésien fournit la recommandation d'une estimation globale – il suppose toutefois de connaître le modèle dynamique). Un bruit d'observation peut être isolé *e.g.* 1) en bénéficiant de **schémas d'échantillonnage répétés** (Fromentin et al. 2001), 2) en étudiant la structure des **résidus** après **lissages** successifs (Theiler and Eubank 1993). On suppose en général qu'un bruit d'observation est "blanc", *i.e.* non corrélé dans le temps, ou soumis à une dynamique rapide (en terme fréquentiel) par rapport à la

dynamique vraie du système, ce qui autorise e.g. les approximations par fermeture (Dieckmann and Law 1999, Law and Dieckmann 1999) ou l'agrégation de variables (Auger et al. 2000). Si les caractéristiques du bruit d'observation (e.g. distribution, amplitude) ne sont pas mesurées ou spécifiées *a priori*, on prend le risque de détruire la dynamique vraie à haute fréquence du système (Wilson and Rand 1997), comme e.g. un phénomène de densité-dépendance directe (Bjornstad et al. 1999). Le problème se complique encore si le bruit est "coloré", i.e. positivement corrélé donc soumis à une dérive. Ce phénomène est particulièrement aigu en géophysique ou en télémétrie, où des biais de grande longueur d'onde dégradent la mesure (Klees et al. 2002). L'écologie marine et l'halieutique y sont aussi très sensibles, puisque les outils d'échantillonnage (e.g. par pêche) sont constamment sujets à modifications.

Séries écologiques. L'étude des trajectoires des systèmes écologiques, si elle est relativement confortable lorsqu'un jeu d'équations est disponible (Nisbet and Gurney 1982), devient donc délicate lorsque l'on travaille sur des séries de données réelles. Leur longueur limitée et le caractère stochastique des systèmes sous-jacents limitent encore l'application de nombreuses méthodes non-linéaires (Casdagli et al. 1991, Eckmann and Ruelle 1992, Kantz and Schreiber 1997, Spagnolo et al. 2004). On cherche alors en général à améliorer par un pré-traitement adéquat la structure du jeu de données initial par des méthodes hybrides de réinterpolation/lissage (Eckmann and Ruelle 1985, Casdagli 1989). Les méthodes existantes d'identification peuvent alors être appliquées directement après ce nettoyage des données (Casdagli et al. 1991, Boudjema and Chau 1996, Boudjema and Cazelles 2001).

Méthodes d'augmentation et d'interpolation

La **linéarisation locale** d'une trajectoire observée sert, on l'a vu, à la mesure de son comportement divergent (exposants de Lyapunov) ou à sa prédiction (Filtre de Kalman Étendu). Elle peut aussi servir au lissage de cette dynamique, afin d'améliorer sa représentation et faciliter la quantification de son déterminisme.

La méthode de prédiction d'Eckmann-Ruelle (Eckmann and Ruelle 1985) repose sur la construction d'un espace des phases empiriques selon la méthodologie de Takens. La nouvelle trajectoire du système est alors un ensemble de points spécifiés par des vecteurs homogènes de taille m . Un système de prédiction linéaire est alors défini pour prédire les coordonnées d'un point X_{i+1} :

$$X_{i+1}=AX_i+B+\varepsilon,$$

où l'on note que A est de dimension $m \times m$ (contre 1×1 pour un modèle AR(1) classique). Ce système est ajusté en considérant les k plus proches **voisins observés** X_j du point X_i dans l'espace des phases (il ne sont pas nécessairement voisins dans le temps). En pratique, on choisit $k=2*m$ pour faciliter l'ajustement du système. Cette méthode particulière d'interpolation permet d'enrichir la trajectoire observée en exploitant l'information dynamique présente dans ses oscillations (Boudjema and Chau 1996). Elle est particulièrement efficace pour des séries courtes.

L'interpolateur dans l'espace des phases peut ne pas être linéaire: si le jeu de données est suffisamment conséquent, il est possible d'ajuster des modèles non-paramétriques, e.g. un réseau de neurones (Boudjema and Cazelles 2001). De telles

méthodes de **renforcement** permettent de valoriser l'information contenue dans les états observés du système dans le **temps** ET dans l'**espace** d'état reconstruit.

Si l'on constate que les méthodes d'identification de la non-linéarité sont souvent basées sur le calcul d'un **taux** de divergence (donc localement sensible), et que les étapes de pré-traitement pour les appliquer reposent sur la construction de **voisinages** (*a priori* plus robustes), on se prend à imaginer une alternative. En effet, il pourrait être possible de définir des statistiques de l'espace des phases reposant directement sur ce concept de voisinage. Elles seraient alors, en principe, moins sensibles à une trajectoire bruitée.

Proposition. Le manuscrit 5 détaille comment une telle méthode peut être construite. L'idée est de profiter de la méthodologie de reconstruction proposée par Takens, puis d'appliquer un filtre sur les voisinages ainsi créés. La représentation des récurrences ("Recurrence Plot" ou RP), introduite en théorie des systèmes dynamiques (Eckmann et al. 1987), serait une méthode candidate.

Les récurrences et les méthodes dérivées ont été principalement développées et appliquées dans le domaine biomédical, afin d'identifier *e.g.* les attracteurs des systèmes ventilatoires et circulatoires (Zbilut et al. 2002), ou des séquences d'acides aminés (Giuliani et al. 2000), en économie (Holyst et al. 2001), en paléoclimatologie (Marwan et al. 2002, Marwan et al. 2003), en physique non-linéaire en général (Ortega and Romanelli 1995, Trulla et al. 1996, Casdagli 1997, Gao and Cai 2000), et de façon anecdotique en écologie marine (Dippner et al. 2002).

Les récurrences ne sont certes pas un concept nouveau: Poincaré décrivait déjà comment les états d'un système dynamique pouvaient visiter des régions confinées de l'espace des phases. La littérature statistique souligne d'ailleurs qu'une récurrence est une relation des plus basiques (Feller 1968). A la différence d'autres méthodes (*e.g.* Fourier, ondelettes...), le calcul des récurrences ne requiert pas de transformation de la données mais seulement leur mise en scène dans un espace des phases empirique.

Récurrences

Sachant une trajectoire de dimension $(1 \times T)$ reconstruite selon les recommandations de Takens:

$$N_{t+1} = f(N_t, N_{t-1}, N_{t-2}, \dots, N_{t-d}) \Leftrightarrow X_i = [N(t), N(t-L), \dots, N(t-(m-1)L)],$$

une matrice de récurrences $(T-L) \times (T-L)$ se construit en renseignant un point (i,j) dès qu'un état X_i est proche d'un autre état X_j (la distance euclidienne $\|X_i - X_j\|$ est alors testée telle que $\|X_i - X_j\| < r$, où r est un seuil dépendant de l'amplitude de la dynamique considérée). Si cette distance est inférieure au seuil, les deux vecteurs sont dits **récurrences** et donc matérialisés par un point de la matrice de récurrence. On notera que la matrice de covariance des vecteurs X est non-paramétrique: ils peuvent représenter des modes oscillatoires, ou tout autre pattern propre à la donnée mesurée. Cette mise en scène de la donnée permet d'apprécier immédiatement la complexité d'une série de mesures biologiques (Webber Jr. and Zbilut 1994). Les récurrences ainsi définies sont en fait des

quasi-récurrences (les récurrences *sensu stricto* étant définies pour les chaînes de Markov).

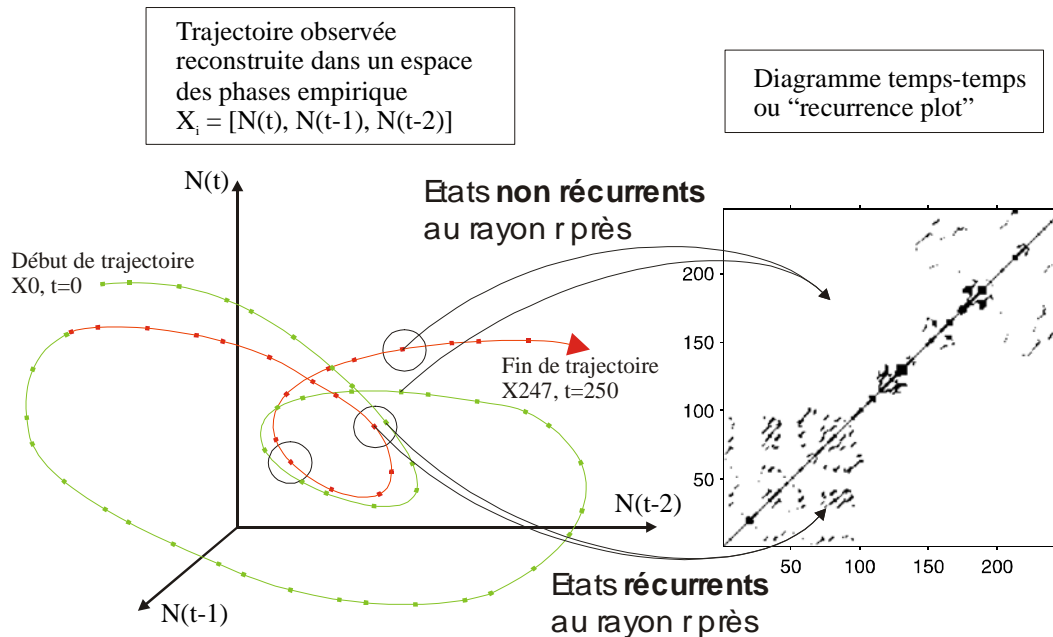


Figure 4-4. Illustration de la construction d'un graphe des récurrences, à partir d'une trajectoire reconstruite dans un espace de Takens

Trois grands types de **patterns** caractérisent les récurrences graphiques:

- (i) les points isolés,
- (ii) les structures diagonales à 45° ou 135°,
- (iii) les structures verticales et horizontales.

Un point isolé signale la proximité de deux vecteurs X_i et X_j : la probabilité est forte qu'elle apparaisse par chance (*i.e.* elle est non significative d'une dynamique). Ainsi un RP (Recurrence Plot) d'un signal purement stochastique (type bruit blanc) se résume à un nuage de points remplissant aléatoirement l'espace (Eckmann et al. 1987). Des structures diagonales à 45° (et parallèles à la diagonale principale) indiquent que pour deux vecteurs X_i et X_j proches, alors X_{i+k} et X_{j+k} sont proches. De façon similaire, les diagonales à 135° indiquent que pour deux vecteurs X_i et X_j proches, alors X_{i+k} et X_{j-k} sont proches. Les structures verticales et horizontales signalent des points de séjour, voire des attracteurs locaux (Gao and Cai 2000), qui piègent le système pendant un temps fini. Ces structures seront de taille relativement égale si la trajectoire possède un mouvement uniforme.

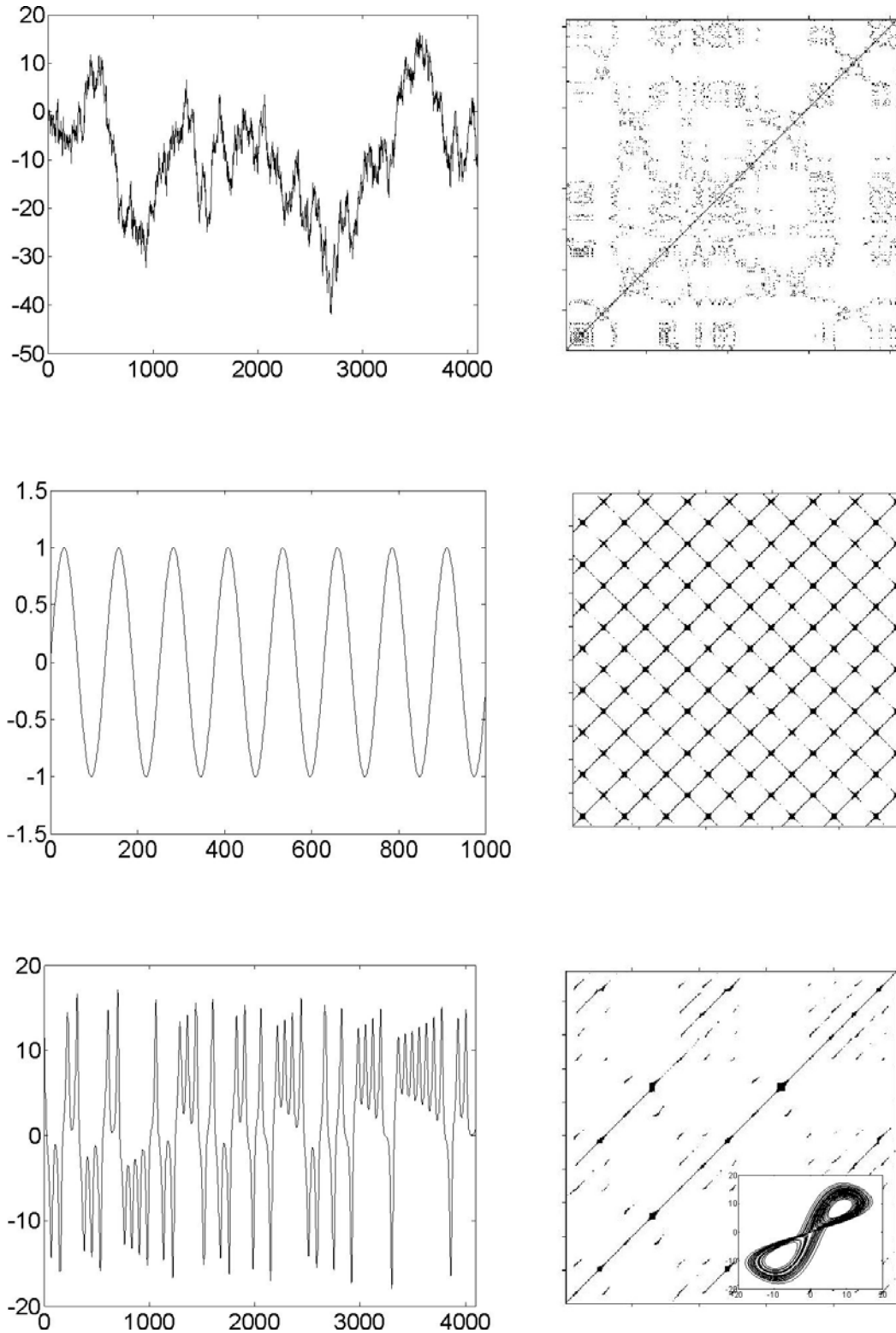


Figure 4-5. Trois exemples de graphe des récurrences dans le cas (a) d'une marche aléatoire, (b) d'une sinusoïde, (c) du système de Lorenz dans le régime chaotique.

La tentation du chaos

Une application intéressante des RPs, outre de pouvoir quantifier expérimentalement la complexité d'une dynamique (e.g. invariants et exposants), est de pouvoir détecter les **séquences de bifurcations** du système (Trulla et al. 1996, Iwanski and Bradley 1998). Le comportement chaotique des modèles écologiques a ainsi reçu une attention particulière: de nombreux auteurs ont identifié la signature d'un chaos déterministe dans des modèles trophiques simples (May and Oster 1976, Hastings and Powell 1991, Schaffer and Kot 1991, Ascioti et al. 1993, McCann and Yodzis 1994, Rinaldi and Solidoro 1998). L'idée que le régime chaotique pouvait expliquer le "paradoxe de la diversité" (Hutchinson 1961) avait de quoi séduire: de récents travaux ont ainsi montré que sous certaines conditions, un tel régime permet de maintenir une grande diversité d'espèces planctoniques dans un milieu globalement homogène (Huisman and Weissing 1999). D'autres investigations théoriques mettent toutefois en doute cette proposition, en établissant que la sélection de stratégies évolutivement stables se traduisait *in fine* par des paramètres de croissance hors du régime chaotique (Joergensen 1997). Il faut enfin noter que les systèmes écologiques ne sont en aucun cas des systèmes fermés, et que l'interaction entre forçages (exogènes) et dynamiques internes (endogènes) complique le problème (Stenseth et al. 2002). En particulier, on notera que les espèces planctoniques interagissent de façon critique avec leur milieu dans son régime turbulent, donc potentiellement chaotique (Takens 1981, Medvinsky et al. 2001)...

La quête même du chaos en écologie reste un objectif controversé (Hastings et al. 1993, Rai and Schaffer 2001). Il apparaît plus réaliste ici d'en détourner les outils, souvent puissants, et de les employer à nos propres fins (celle de l'identification des mécanismes écologiques). **Changements de régime**, forme et linéarité des **relations fonctionnelles**, périodicité des **oscillations** et **interaction signal-bruit** sont autant de phénomènes écologiques d'intérêt pouvant être révélés et quantifiés par la méthode des récurrences.

La représentation graphique des récurrences d'un signal, bien que très esthétique (Gao and Cai 2000), est insuffisante pour l'identification ou la classification des mécanismes sous-jacents. La **quantification statistique** de ces récurrences est non seulement possible (un RP étant généralement composé d'un grand nombre de points), mais surtout très indiquée (Zbilut et al. 1998, Marwan et al. 2002, Zbilut et al. 2002). Les statistiques proposées par ces auteurs sont par exemple: (i) le pourcentage REC de points récurrents couvrant la matrice, (ii) le déterminisme DET de la trajectoire comme le pourcentage de points formant des structures diagonales, (iii) la complexité de ce déterminisme, comme l'entropie EDL (au sens de Shannon) de la distribution de ces lignes diagonales, (iv) la longueur de la plus longue diagonale LDL (son inverse étant directement reliée au plus grand exposant de Lyapunov positif – Gao et Cai 2000), (v) la laminarité LAM comme le pourcentage de points récurrents formant des structures verticales ou horizontales, (vi) le temps moyen de piégeage TT comme la longueur moyenne de ces structures...

Le calcul seul de ces statistiques reste toutefois peu informatif dans l'analyse exploratoire des systèmes dynamiques (Theiler et al. 1992, Zbilut and Webber Jr. 1992, Webber Jr. and Zbilut 1994, Barnard et al. 2001).

Leur **interprétation** correcte relève, encore une fois, du test d'hypothèse, *i.e.* de la construction d'une hypothèse H_0 et d'un modèle nul. En d'autres termes, "la dynamique que j'observe est-elle plus structurée, plus organisée que ce que j'observerais dans un monde aléatoire?". Tester ceci revient à construire l'**homologue statistique** (ou "**surrogate**" en langue de Shakespeare) de la série mesurée, évaluer la statistique de test appropriée sur les deux séries, puis comparer ces deux valeurs. Le problème de la **significativité** peut être résolu en générant, par simulation de Monte-Carlo, un **ensemble** de ces homologues statistiques (Allen and Robertson 1996, Schreiber and Schmitz 1996). La problématique se ramène alors à la construction et au test d'une hypothèse nulle. Or on rappelle qu'il est possible de générer, par des méthodes de bootstrap (cf. chapitre 2.3), des séries aléatoires de différentes propriétés. En les associant chacune à un modèle nul, ces méthodes de simulation conditionnelle permettent de formuler différents tests d'hypothèse (article 5).

Courte introduction au test d'hypothèse dynamique

Les trois algorithmes de rééchantillonnage introduits en au chapitre 2.3 sont présentés ici à des fins de test d'hypothèse, ou "surrogate testing" (Theiler et al. 1992, Barnard et al. 2001).

Ho: BRUIT BLANC

Le rééchantillonnage par bootstrap d'une série de données détruit toute structure non-linéaire, blanchit le spectre (*i.e.* $\Phi(\omega)=\text{constante}$) et ne conserve que la distribution temporelle (ou spatiale) initiale. En d'autres termes, le modèle nul sous-jacent est: $X_t = \xi_t$. On teste donc la possibilité d'un système stochastique, dont les réalisations sont totalement décorréllées dans le temps (ou l'espace). En théorie, l'essentiel des résultats asymptotiques en théorie de l'estimation sont développés pour de tels systèmes. En pratique, bien peu de processus géophysiques sont capables de produire de telles sorties (Donoho 1992). Un tel test s'applique donc préférentiellement aux résidus d'un ajustement ou à un processus d'observation plutôt qu'à une mesure brute. En écologie, si un tel modèle permet *a priori* de quantifier le degré de structure d'une série de données (Dippner et al. 2002), on notera qu'il rend compte d'une population parfaitement régulée de façon endogène, ce qui est de peu d'intérêt pour notre propos. Le problème de tester une telle hypothèse sera développé dans la suite.

Ho: BRUIT ROUGE

Le bootstrap des phases du signal (dans le domaine de Fourier) permet de générer des séries de même spectre que l'original. En d'autres termes, la fonction d'autocorrélation est conservée mais toute autre dynamique non-linéaire est détruite, ce qui revient à tester le modèle suivant:

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-d}) + \varepsilon_t.$$

Ce modèle a un intérêt particulier en géophysique puisque la majorité des mesures des systèmes naturels présente un spectre rouge, ou localement en $1/f^\alpha$, où $0.5 < \alpha < 1.5$. On parle alors de "bruit en $1/f$ ", ou "bruit de Flicker", qui permet de caractériser de nombreux phénomènes turbulents (voir *e.g.* les travaux de Kolmogorov), la fréquence d'évènements sismiques (Mandelbrot and Wallis 1969) ou océanographiques (Press 1978, Steele 1989). On se référera aux travaux de Milotti (2001) pour une revue complète de ces signaux. De tels spectres peuvent ainsi révéler l'existence d'une invariance d'échelle, d'un continuum de mécanismes dans l'espace ou le temps, de cascades de processus stochastiques (Milotti 1995, 2001)... Leur quête et leur interprétation en écologie est le sujet d'un débat animé, et ne sera pas abordé ici (Gisiger 2001). Il est suffisant de rappeler que le modèle précédent est fondamentalement linéaire, et correspond à une population où la densité-dépendance est quasi-inexistante, ou de nature statistique seulement (Saitoh et al. 1997). Ainsi la forme $X_t = f(X_{t-1}) + \varepsilon_t$ correspond à une population non régulée (*i.e.* de variance toujours croissante) et constitue un modèle nul intéressant (celui d'un bruit filtré) pour l'enquête écologique (Jonzén et al. 2002, Jiang and Shao 2004).

Ho: BRUIT ROUGE TRANSFORME

L'algorithme itératif de rééchantillonnage des phases permet de préserver les distributions fréquentielle ET temporelle du signal: si le modèle sous-jacent correspond toujours à un processus autorégressif linéaire, il est ici sujet à un processus d'observation stationnaire mais non-linéaire (*i.e.* c'est une non-linéarité statique et sans mémoire).

$$X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-d}) + \varepsilon_t$$

$$Y_t = g(X_t)$$

En d'autres termes, le modèle nul est un "state-space model" nul, ainsi que proposé par Takens (1981). Il appartient alors à l'expérimentateur de préciser la forme de la relation d'observation, grâce *e.g.* à des échantillons multiples (Fromentin et al. 2001).

Ces homologues statistiques, ou "surrogates", sont d'abord et avant tout **linéaires** et **stochastiques**. Le rejet de l'hypothèse nulle correspondante n'implique pas forcément le rejet du modèle exact, mais d'abord la présence de relations non-linéaires et/ou déterministes dans les données. On doit distinguer les **hypothèses nulles simples** (où les paramètres du modèle nul sont précisés, e.g. ajustés à 0) des **hypothèses nulles composites** (où l'on se réfère à une famille de processus). Si la distribution de la statistique discriminante dépend des paramètres du modèle nul, cette statistique est dite **pivotal** (Theiler et al. 1992). Des questions qui nécessiteraient autrement la connaissance du modèle sous-jacent peuvent être alors adressées aux données:

- (i) Quel est le degré de stochasticité du système?
- (ii) Quel est son degré de linéarité (de non-linéarité)?
- (iii) Est-il stationnaire ou non-stationnaire (*i.e.* synthétisable par un ensemble invariant de dimensions)?
- (iv) Quelle est la dimension de son (ou ses) attracteur(s), s'ils existent?
- (v) Peut-il générer des événements remarquables?

Le manuscrit 5 fournit l'exemple d'une application contrôlée. Cet exemple met à profit les cycles générés par un simple modèle de type NPZ, où les fluctuations du compartiment P sont observées avec un bruit d'observation de grande variance (de l'ordre de la variance générée par le processus lui-même, *i.e.* un cas de modélisation difficile). Si les cycles ont une signature spectrale bien visible, leur significativité n'est pas évidente. La construction d'une séquence de récurrences graphiques sur la série observée permet de révéler ces patterns périodiques, puis de tester leur caractère déterministe (*i.e.* leur régularité). La distribution des observations bruitées étant quasi-gaussienne, l'hypothèse H_0 est celle d'un bruit rouge (*i.e.* un modèle caractérisé seulement par sa distribution fréquentielle, son autocorrélation, ou de façon équivalente un jeu de paramètres d'autorégression). Un tel "surrogate", ou homologue stochastique de même spectre est fourni en exemple: s'il présente effectivement des variations de même période que l'original, l'organisation de ces cycles (visibles sur le RP correspondant) ne montre aucune structure particulière. La simulation d'un ensemble de ces "surrogates" permet alors le test des statistiques de récurrence. On montre ainsi que la série observée présente un surplus de patterns récurrents par rapport au modèle nul: on rejette l'hypothèse nulle H_0 . La modélisation déterministe de ces cycles est donc indiquée.

La méthode a ensuite été appliquée aux séries de capture des madraques siciliennes et sardes (Ravier 2003). Ces séries étant de longueur relativement limitée (130 à 180 points consécutifs), les résultats doivent être interprétés avec précaution. La dimension de l'espace de reconstruction a ainsi été estimée à 6-8 par la méthode des faux voisins, et fixée à 8 pour plus de marge (Kantz and Schreiber 1997). Le phénomène échantillonné étant lié à un comportement migratoire (lié à la reproduction et *a priori* annuel), la résolution d'échantillonnage a été fixée

à 1 an, en accord avec la fonction d'information mutuelle des séries. Sans surprise, l'hypothèse nulle d'un bruit blanc (des fluctuations décorréées) a été rejetée au seuil de 95%, l'essentiel des statistiques de récurrence reposant en dehors des distributions nulles. L'hypothèse nulle d'un bruit rouge simple a elle aussi été rejetée, de même que, dans une moindre mesure, l'hypothèse nulle d'un bruit rouge transformé: quatre des six séries montrent ainsi un surplus de patterns récurrents par rapport à leurs homologues stochastiques. Ce dernier résultat indiquerait que, dans la plupart des cas, il est insuffisant d'employer un simple modèle linéaire autorégressif transformé. Il existe en particulier une non-linéarité **dynamique**, et non statique, *i.e.* ce qui serait attendu dans le cas d'un processus d'observation extrinsèque au système (Takens 1981).

Quelle doit être l'étape suivante de l'enquête? Des fluctuations bornées, l'existence de patterns récurrents sont sans doute la signature de phénomènes de **régulation** dans la population de thon rouge. Les principes de la dynamique des populations (Berryman, 2003) supposent que ces phénomènes régulateurs agissent avec peu ou pas de délai temporel, donc à haute fréquence. Une autre hypothèse fait intervenir des fluctuations exogènes: si la population de *Thunnus thynnus* présente une réponse (linéaire ou non, avec ou sans délai) à ces **forçages**, il est possible que les caractéristiques déterministes du signal ne soient que la signature de quelque variable géophysique. Le climat est bien évidemment le premier suspect (Ravier and Fromentin 2004), d'autant plus, on l'a vu, que le signal océanique présente un spectre rouge, dominé par les basses fréquences (Steele 1985, Vasseur and Yodzis 2004). La plage entière de fréquences est donc candidate pour la suite. Comment démêler les origines endogènes et exogènes du signal, *i.e.* séparer les domaines fréquentiels d'action des différents processus?

Une approche exploratoire originale est proposée pour étudier la dynamique temporelle d'une série d'observation. La construction d'un espace des phases empiriques, combinée à l'analyse des états récurrents et à l'emploi "d'homologues stochastiques", ouvre une voie intéressante pour le test d'hypothèse dynamique.

4.3. Retour aux techniques classiques

Les méthodes **exploratoires** dérivées de l'étude du chaos déterministe reposent sur des hypothèses minimalistes. Or ces outils, s'ils permettent d'aborder des questions générales importantes, sont encore peu développés en écologie. Le besoin de dynamiques convenablement cartographiées et fournies limitent souvent leur application aux modèles quasi-déterministes et à la recherche de comportements chaotiques dans les modèles de dimension raisonnable. Lorsqu'il s'agit de données réelles,

l'analyste dispose d'autres outils permettant d'extraire et décomposer la dynamique observée, sous réserve de certaines hypothèses. En particulier, l'essentiel des techniques classiques en séries temporelles s'offrent à lui sous réserve de **linéarité** ou de **stationnarité** de son problème. On montre dans la suite comment l'emploi de ces techniques peuvent enrichir nos connaissances sur la dynamique de *Thunnus thynnus*.

Décomposition fréquentielle. On entend ici par décomposition toute projection de la dynamique $X(t)$ sur un nouvel ensemble d'axes. Ces axes peuvent être matérialisés par des fonctions paramétriques (e.g. séries de Fourier, polynômes de Legendre) ou non-paramétriques (e.g. Empirical Orthogonal Functions, formulation avec délais). Les méthodes de Fourier offrent l'avantage d'être facilement interprétables, et de profiter d'un grand nombre de résultats reliant les domaines temporel et fréquentiel.

Relation entre l'autocorrélation et le spectre de variance

La fonction d'**autocorrélation** (ACF) est un outil de diagnostic largement employé en séries temporelles (Chatfield 1999), et permet de révéler objectivement la période d'un phénomène cyclique. La "mémoire" du système peut ainsi être caractérisée par la décroissance de cette fonction: un processus de Markov, qui n'a d'autre mémoire que celle définie par son état présent, présentera des fluctuations décorrélatées. L'ACF est généralement utilisée à des fins de modélisation, pour e.g. estimer le modèle autorégressif correspondant selon la méthode de Yule-Walker (Walker 1931).

$$c(\tau) = \langle x(t)x(t+\tau) \rangle \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau) dt$$

Le spectre de variance organise cette information en terme de fréquence et non de délai: c'est en fait la transformée de Fourier de la fonction d'autocorrélation. Le spectre de variance est utilisé en analyse exploratoire (Legendre and Legendre 1998, Dale et al. 2002).

$$P(\omega) = 2 \int_0^{\infty} c(t) \cos(\omega t) dt$$

Grâce à des méthodes combinées d'analyse en composantes principales et en séries de Fourier, Ravier et Fromentin (2001) ont montré comment les longues séries de capture (>80 ans) partageaient la même structure fréquentielle à moyen terme (i.e. un pic de variance pour des périodes de 15-30 ans). Ces auteurs ont aussi noté comment les cycles de grande longueur d'onde (i.e. de période 100-120 ans) se comportaient comme des tendances hautement significatives (50-70% de variance expliquée) à l'échelle des séries (de taille ~80-180 ans).

Si la variabilité à plus haute fréquence (de période < 10 ans) est significative, elle apparaît secondaire par rapport aux pseudo-cycles de grande période. Au vu de la synchronie à long terme des madragues, et de la difficulté à reproduire les cycles de grande longueur d'onde à l'aide de simples modèles autorégressifs (Ravier 2003), il est naturel de reconnaître ces fluctuations à long terme comme la caractéristique la plus marquante de la population.

Or la fonction d'autocorrélation vise à fournir un renseignement précieux sur **l'échelle de temps caractéristique** (*i.e.* la constante de temps) du système. Si la dynamique mesurée est caractérisée par des fluctuations de grande période, on dit que le système présente une "mémoire" à long terme. En d'autres termes, la fonction d'autocorrélation sera positive sur une grande plage de délais, avec un premier passage à zéro tardif (de façon plus critique encore si la série est non-stationnaire) (Chilès and Delfiner 1999). Réduire le système à une équation **continue** basée sur cette seule constante de temps (*e.g.* une équation de Langevin) revient donc à fortement pondérer la dynamique à basse fréquence, en négligeant la plus haute fréquence. Enfin, dans le cas de dynamiques **discrètes** (*e.g.* des mesures annuelles d'abondance), la fonction d'autocorrélation a des difficultés à capturer **l'ordre** du modèle autorégressif correspondant (Chatfield 1999), soit la constante d de l'écriture $X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_d X_{t-d} + \varepsilon_t$. Suivant Berryman et Turchin (2001), deux variantes sont donc introduites dans la suite pour étudier la structure des séries de capture: **l'autocorrélation partielle (PACF)**, et **l'autocorrélation partielle du taux d'accroissement (PRCF)**.

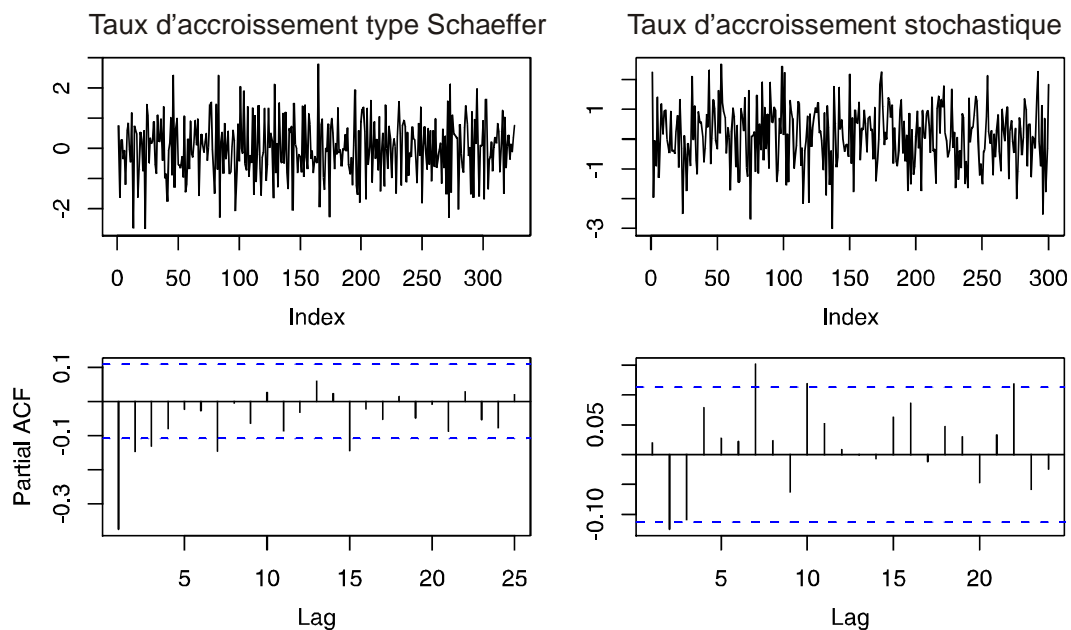


Figure 4-6. Autocorrélation partielle du taux d'accroissement d'un modèle de type logistique (à gauche), et du taux d'accroissement d'une population non compensée (*i.e.* une marche aléatoire, à droite)

Autocorrélation partielle et hypothèse nulle

Différents auteurs (Berryman and Turchin 2001) ont montré que la fonction d'autocorrélation ne permettait pas d'estimer de façon rigoureuse l'ordre d'un processus autorégressif (*i.e.* le délai de réponse d'un phénomène de régulation). Ceci est dû au fait que **l'inertie** naturelle du système masque les processus de **régulation**.

La fonction d'**autocorrélation partielle** (Partial Autocorrelation Function - PACF) a été proposée afin de quantifier pas-à-pas la contribution de chaque délai, en fixant les effets des délais précédents.

$$PACF(m) = \begin{cases} ACF(1) & \text{pour } (m=1) \\ \frac{ACF(m) - \sum_{j=1}^{m-1} PACF(m-1, j)ACF(m-1)}{1 - \sum_{j=1}^{m-1} PACF(m-1, j)ACF(j)} & \text{pour } (m>1) \end{cases}$$

Il est alors possible d'estimer l'ordre d du modèle autorégressif sous-jacent:

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_d X_{t-d} + \varepsilon_t,$$

qui constitue le modèle nul du test basé sur la PACF. Les valeurs de la PACF pour des délais supérieurs à d seront statistiquement non-différents de zéro. Un tel modèle nul est malheureusement peu satisfaisant en écologie: les variations d'abondance d'une populations étant mesurées en nombre d'individus, il est nécessaire d'écrire la dynamique sous la forme suivante:

$$X_t = X_{t-1} R(X_{t-1}, X_{t-2}, \dots, X_{t-d}, \varepsilon_t),$$

où $R(\dots)$ définit le taux de variation *per capita* de la population. En définissant $R_t = \ln(X_t/X_{t-1})$ et en travaillant dans la suite sur les logarithmes, soit $Z = \ln(X)$, le nouveau modèle testé s'écrit:

$$R_t = Z_t - Z_{t-1} = a_0 + a_1 Z_{t-1} + a_2 Z_{t-2} + \dots + a_d Z_{t-d} + \varepsilon_t.$$

L'ordre de ce modèle peut alors être évalué en étudiant **l'autocorrélation partielle du taux d'accroissement** (Partial Rate Autocorrelation Function –PRCF).

Il est important de rappeler **le modèle nul** sous-jacent à ces deux écritures. Pour la PACF, il s'agit de tester (1) $X_t = a_0 + \varepsilon_t$, soit $a_1 = a_2 = \dots = a_d = 0$, donc une population décrivant des fluctuations aléatoires non corrélées autour de sa moyenne a_0 (*i.e.* un bruit blanc). Pour la PRCF, ce modèle nul est (2) $R_t = Z_t - Z_{t-1} = a_0 + \varepsilon_t \Leftrightarrow Z_t = a_0 + Z_{t-1} + \varepsilon_t$ soit une population décrivant une **marche aléatoire** (*i.e.* un mouvement brownien, ou bruit brun). Ces deux modèles nuls sont fondamentalement différents. Le modèle (1) décrit une population **parfaitement régulée** (un bruit blanc ou IID étant contraint en premier lieu par sa moyenne). Un tel modèle nul est inintéressant voire handicapant pour des questions écologiques, et doit donc être écarté au profit de l'hypothèse d'une population **non régulée** (modèle 2). Ceci explique pourquoi certains auteurs ont pu détecter un phénomène de régulation significatif sur une série de type IID ou bruit blanc, mettant en cause l'efficacité du test par PRCF (Berryman and Turchin 2001).

MORALITE

La PACF fournit un outil précieux de diagnostic en dynamique des populations, mais elle n'est associée à un modèle nul satisfaisant qu'une fois appliquée au taux d'accroissement (*i.e.* une différence de premier ordre).

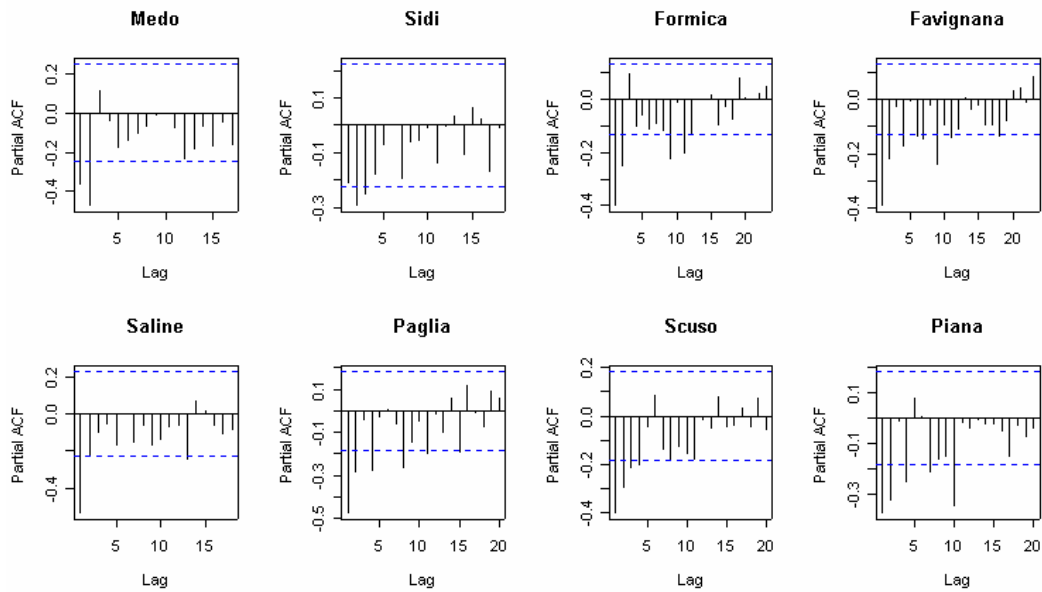


Figure 4-7. Autocorrélation partielle du taux d'accroissement des séries des madrugues méditerranéennes. Les variations des enveloppes de significativité sont dues à l'emploi du critère de Bartlett, dépendant de la longueur de la série.

La PRCF des séries de notre étude apparaît significativement différente de zéro pour les délais 1 et 2. C'est la signature d'un phénomène de **régulation**, ou au moins d'un taux d'accroissement (*i.e.* une dérivée de premier ordre) variable. Un modèle candidat pour approcher la dynamique des séries de capture serait donc:

$$X_t = X_{t-1}R(X_{t-1}, X_{t-2}, \varepsilon_t)$$

Avant d'entamer la construction d'une nouvelle hypothèse (mécaniste et explicative cette fois, et non plus d'ordre statistique), on notera que la contribution de chacun de ces délais (en particulier le lag 2) apparaît inégale selon la série. De plus, le critère de Bartlett employé pour tester la significativité est dépendant du nombre d'observations disponibles, et produit donc des résultats non homogènes. Peut-on améliorer l'estimation de ce processus agissant à haute fréquence? Dans le même esprit, l'estimation des fluctuations à basse fréquence pose problème. Si leur période peut être relativement bien estimée sur la base des plus longues séries, elle est égale ou supérieur à la taille de la plupart des séries. Est-il possible d'estimer la fréquence de ces oscillations en profitant de l'ensemble des observations (soit 327 ans de 1634 à 1960)?

Méta-analyse. La construction d'indices synthétiques, ou l'extraction d'une information commune à plusieurs processus d'observation fait appel aux méthodes de **méta-analyse**. Le niveau organisationnel de cette information est en général incrémenté lors de la méta-analyse (*e.g.* sous-population → population, populations → espèce, etc...). Cette approche a reçu un intérêt particulier en halieutique et surtout en gestion des stocks, où les séries de données sont rares et souvent de petite taille (Myers and Mertz 1998, Myers et al. 2001). L'estimation correcte de paramètres d'importance (*e.g.* taux de croissance,

coefficients de dépensation, survie des premiers âges...) est nécessaire pour étudier les interactions entre traits d'histoire de vie de l'espèce, activité humaine et environnement. Or l'ajustement de modèles de production, même simples, nécessite souvent des jeux de données fournis et de grande taille, afin de minimiser l'effet de la variance d'observation, ou estimer des paramètres de non-linéarité sensibles (e.g. exposants). L'ajustement d'un modèle agrégé, traitant conjointement les mesures de différentes stations d'observation, permet alors de compenser ces problèmes (Gibson and Myers 2003). Notre problème n'est pas *sensu stricto* l'estimation de paramètres conjoints, mais l'extraction d'une information commune à toutes les madragues. En d'autres termes, on cherche à estimer le spectre fréquentiel global des variations de capture (i.e. on cherche à modéliser l'information contenue dans la donnée plutôt que la donnée elle-même). Il est donc nécessaire d'employer une méthode méta-analytique préservant au mieux les propriétés dynamiques (fréquentielles) de la variable d'état. Trois propositions sont faites ici pour estimer une série synthétique des fluctuations de captures, complétant la série construite par filtre médian par Ravier et Fromentin (2001). On cherche en particulier à retenir la méthode préservant l'intégralité du spectre fréquentiel des captures (hautes-basses fréquences). La qualité de la reconstruction n'est pas appréciée au sens des moindres carrés ou grâce à un critère d'information (les méthodes étant différentes), mais de façon qualitative par le spectre de variance et la fonction d'autocorrélation partielle. On emploie donc ces deux outils comme des "sondes", *sensu* Turchin (2003).

La première proposition fait intervenir la résolution d'un **système de krigeage**, basé sur l'autocorrélation empirique de la variable observée (Matheron 1971). L'estimation de la fonction d'autocorrélation, s'il est délicat lorsque la moyenne spatiale est inconnue ou que les mesures sont irrégulièrement espacées, peut s'effectuer en estimant le semi-variogramme empirique (Cressie 1993, Chilès and Delfiner 1999). En d'autres termes, employer le variogramme permettrait *a priori* de conserver le spectre de variance empirique.

Krigeage temporel: une forme de méta-analyse

L'estimation d'une dynamique temporelle commune à plusieurs stations peut s'écrire sous la forme d'un système de krigeage:

$$X^*_{t_0} = m + \sum_{\alpha=1}^n \lambda_{\alpha} (X_{\alpha} - m) \Leftrightarrow \sum_{\beta=1}^n \lambda_{\beta} C_{\beta\alpha} = C_{\alpha 0},$$

où $C(x_{\alpha} - x_{\beta}) = C_{\alpha\beta}$ est la fonction de covariance concernant deux points X_{α} et X_{β} séparés dans l'espace et le temps. La résolution de ce système nécessite de posséder une représentation de cette covariance. En pratique, on calcule le semi-variogramme (spatial ou temporel) du jeu de données:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} (Z(X_i) - Z(X_j))^2,$$

puis on modélise ce variogramme empirique par une fonction paramétrique adéquate (Cressie 1993). Différents modèles, en fonction des propriétés de la donnée, peuvent être envisagés (e.g. modèle sphérique, exponentiel, à trou, sinusoidal...).

Le semi-variogramme empirique des neuf séries de capture est caractéristique d'une dynamique cyclique avec effet d'observation (ou de pépité): celui-ci a été construit en postulant que le même processus était mesuré au temps t par chaque madrague. La seule dimension du variogramme est donc le temps. L'ajustement d'un modèle "en vague" et à effet de pépité est à première vue satisfaisant, mais la reconstruction par krigeage qui en découle est définitivement trop lisse: les trois pseudo-cycles de période ~ 100 ans sont clairement visibles, mais aucune dynamique à plus haute fréquence n'est extraite. L'estimation des cycles à 15-30 ans (Ravier and Fromentin 2001) n'apparaît possible qu'au prix de l'ajustement d'un modèle à palier et à effet de pépité limité, beaucoup moins vraisemblable au vu du variogramme empirique. Il semble, au vu de ces résultats, que l'emploi d'une fonction de covariance à des fins de reconstruction accorde trop de poids à l'information présente dans les basses fréquences de la donnée. En pratique, il serait nécessaire de considérer e.g. un variogramme gigogne, avec un effet de pépité limité et précisé *a priori*. Cet effet de pépité (un analogue du rapport signal-bruit aux hautes fréquences) devrait donc être précisé au préalable pour assurer une reconstruction satisfaisante.

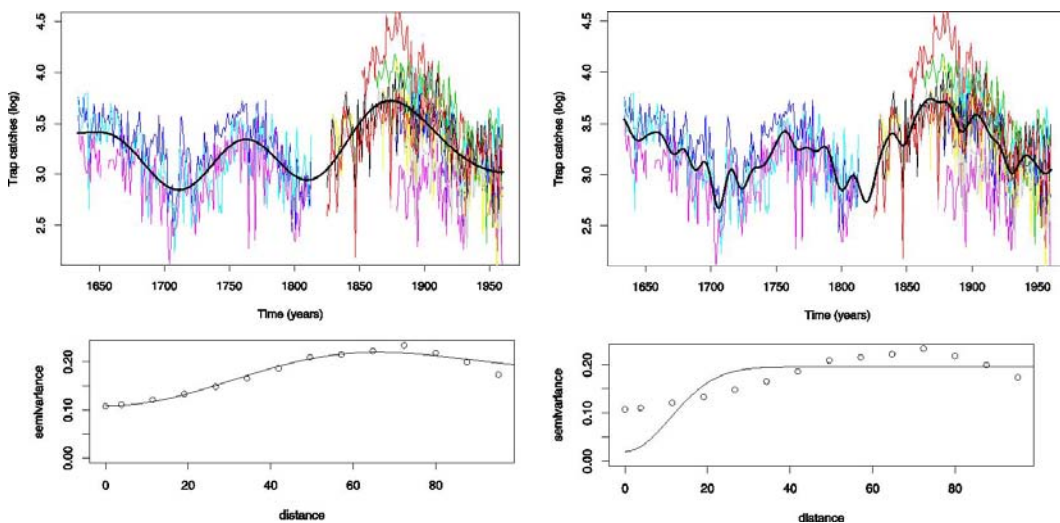


Figure 4-8. Krigeage des séries temporelles de captures basées sur un modèle variographique "en vague" (à gauche), et un modèle à palier forcé (à droite).

La deuxième proposition fait intervenir l'ajustement Bayésien d'un modèle simple de dynamique des populations, en tenant compte de plusieurs processus d'observations. C'est en fait un "state-space model" à équations d'observation multiples. La dynamique f doit donc être explicitée: les données deviennent ici limitantes. En effet, elles ne sont structurées ni en âge ni en taille, car archivées sous des formes diverses au cours des siècles. Le fait de disposer de seuls indices de biomasse (en fait des captures) limite le choix aux modèles agrégés. On emploiera ici, pour illustration, un simple modèle de Schaeffer.

Code BUGS pour un modèle de Schaeffer à observations multiples

Le code ci-après décrit comment un modèle simple (de type Schaeffer) peut être ajusté aux 9 séries (incomplètes) de captures de madragues sur la période 1634-1960. Sa simplicité apparente est liée au grand degré de compactage du langage BUGS (Spiegelhalter et al. 1996), et à l'encapsulation de la plupart des étapes complexes d'échantillonnage.

```

model Schaefer; {
  ## Spécification du prior
  r~dlnorm(1,0.10)*
  K~dnorm(3.5,10)
  tau~dnorm(100,100) # Prior positif pour l'erreur de processus
  for (i in 1:nt) { I[i]~dnorm(1,10)      # Efficacités des engins
                  Ytau[i]~dnorm(10,0.1) # Erreurs d'observation }
  B[1]~dnorm(3,10)

  ## Itérations
  for (t in 2:npts) {
    # Propagation du vecteur d'état
    Bmean[t-1]<-B[t-1]+r*B[t-1]*(1-B[t-1]/K)
    B[t]~dnorm(Bmean[t-1],tau)
    for (i in 1:nt) { # Processus d'observation (un par madrague)
      n[i,t]<-I[i]*B[t]
      Y[i,t]~dnorm(n[i,t],Ytau[i])    }
    }
  }
#end
}

```

* sous WinBUGS, le second terme est une précision=1/écart-type

L'ajustement de ce modèle donne des résultats à première vue plus satisfaisants qu'un krigeage: la série synthétique ainsi créée est caractérisée par un spectre de variance maximale à 110-160 ans, et une pente spectrale de -1.65 contre -3 , donc plus homogène (*i.e.* accordant plus de variance aux hautes fréquences que le krigeage précédent). Toutefois la fonction d'autocorrélation partielle, appliquée au taux d'accroissement de la variable d'état X , ne montre plus de patterns caractéristiques (haute fréquence). Comment expliquer cette différence? L'hypothèse la plus vraisemblable est que le signal haute fréquence ait été attribué au processus d'observation (écrit comme un bruit blanc, donc substantiellement caractérisé par de hautes fréquences). Une autre possibilité est que l'échantillonneur de Gibbs employé dans le logiciel accorde naturellement un poids important à l'autocorrélation des données, et détruise en partie leur dynamique haute fréquence.

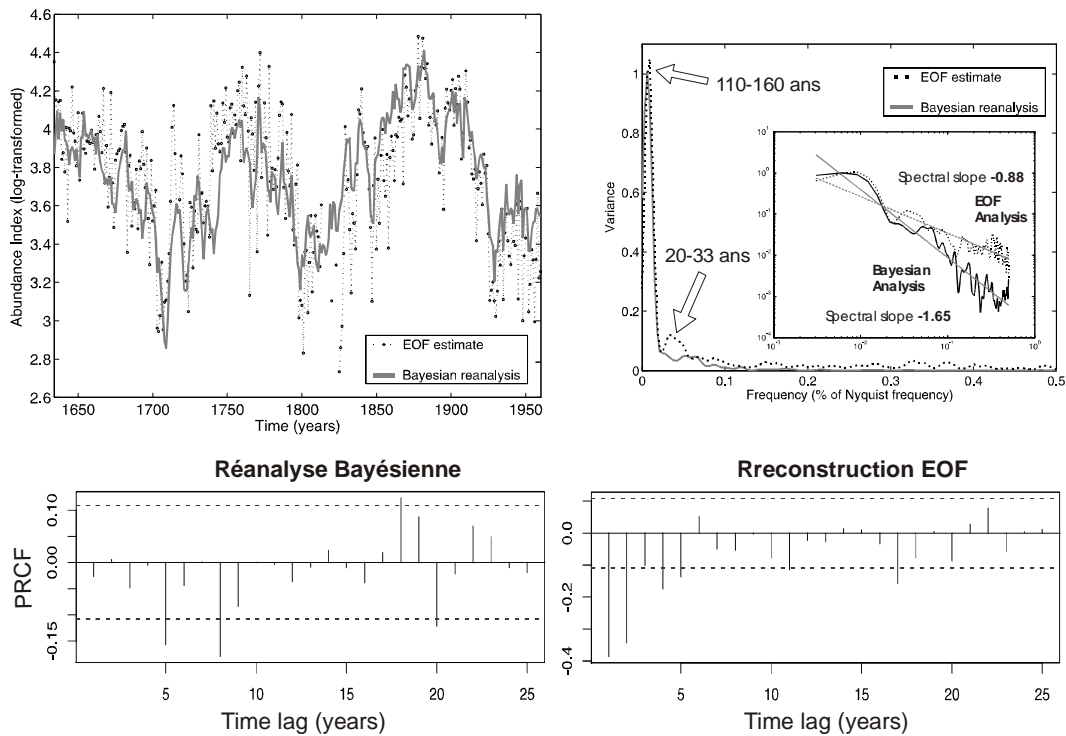


Figure 4-9. Comparaison de la reconstruction par ajustement bayésien d'un modèle de Schaefer à observations multiple, et par fonctions orthogonales empiriques (EOFs). Le spectre de variance de cette dernière apparaît clairement plus homogène, avec une mise en valeur du pseudo-cycle intermédiaire à 20-33 ans. Le pattern de densité-dépendance est clairement significatif aux lags 1 et 2.

La troisième proposition fait appel à la méthode d'analyse objective présentée au chapitre 2, détournée et appliquée au domaine temporel comme l'estimateur de krigeage. La décomposition des séries en EOF (analogue à l'analyse en composantes principales, ou ACP) est employée pour ré-estimer les valeurs manquantes. Les axes étant définis de manière empirique, on s'attend à ce qu'une telle méthode opère sur l'ensemble du spectre fréquentiel (*i.e.* hautes et basses fréquences partagées par les stations d'observation doivent être *a priori* reconstruites). Cette méthode a permis de générer une série synthétique plus homogène en terme fréquentiel, et plus proche des données initiales (*i.e.* de pente plus vraisemblable: -0,88 contre -1,65 pour la reconstruction Bayésienne). En particulier, le cycle de 20-33 ans déjà mis en évidence dans les précédents travaux (Ravier and Fromentin 2001) apparaît clairement visible. Une telle série pourrait apparaître qualitativement moins satisfaisante (car relativement bruitée). Pourtant cette dynamique haute fréquence est nettement structurée: l'autocorrélation partielle du taux d'accroissement est négative et significative sur deux délais successifs. Ceci constitue donc bien un signal haute fréquence partagé par les séries de madragues, et non un bruit d'observation. Il apparaît donc hasardeux d'employer les techniques basées sur l'autocorrélation (*e.g.* chaînes de Markov, variogramme) pour le mettre en évidence de façon aveugle. L'autocorrélation, aspect fondamental de l'analyse exploratoire, joue en notre défaveur pour les problèmes d'estimation. La reconstruction par

EOF, si elle reste d'ordre expérimental, apparaît beaucoup plus satisfaisant puisque l'ensemble du spectre semble préservé, car cette technique extrait l'information en fonction de la variance commune aux stations, quelle que soit son domaine fréquentiel. La pente spectrale de la série synthétique apparaît plus faible, témoignant d'un meilleur ratio entre hautes et basses fréquences de la dynamique commune aux madragues. La présente étude pourrait être approfondie, en envisageant des raffinements tels que: la pondération de l'effet de chaque madrague par une distance géographique, l'ajustement de modèles autorégressifs type Gompertz (Royama 1992), etc...

Dans la suite, l'intérêt est reporté sur la réponse fréquentielle des populations: on s'interroge en particulier sur l'origine des fluctuations à long terme de la population (endogène contre exogène), et les moyens nécessaires à son identification et son évaluation.

Analyse fréquentielle. Une approche relativement récente en écologie quantitative est l'analyse fréquentielle des modèles de dynamique des populations (Kaitala et al. 1997, Laakso et al. 2001, McCann et al. 2003). Si l'interaction trait d'histoire de vie/environnement est maintenant reconnue comme étant un problème central (Steele and Henderson 1984, Spencer and Collie 1997), sa démonstration empirique ou théorique n'est pas triviale (McGowan et al. 1998, Bjornstad et al. 1999). Ainsi la réponse passive d'une population à l'environnement (par un ensemble de processus linéaires) pourra être mise en évidence facilement par la mise en corrélation. L'interaction entre environnement et phénomènes densité-dépendants sera, elle, plus délicate à extraire et nécessite souvent une étude théorique préalable (Gurney et al. 2001, Jonzén et al. 2002, McCann et al. 2003).

Principe de l'analyse fréquentielle

De nombreux systèmes géophysiques et biologiques regroupent des composants interagissant à diverses échelles spatiales et temporelles (e.g. le couple océan-atmosphère). Il est possible d'aborder de tels problèmes par l'agrégation de variables et l'analyse de dynamiques lentes/rapides (Auger et al. 2000). Une illustration simplifiée peut être fournie par l'équation de Langevin, où un terme de forçage x impose une dynamique rapide au terme y , de dynamique plus lente:

$$\dot{y} = -\alpha y + \sigma x$$

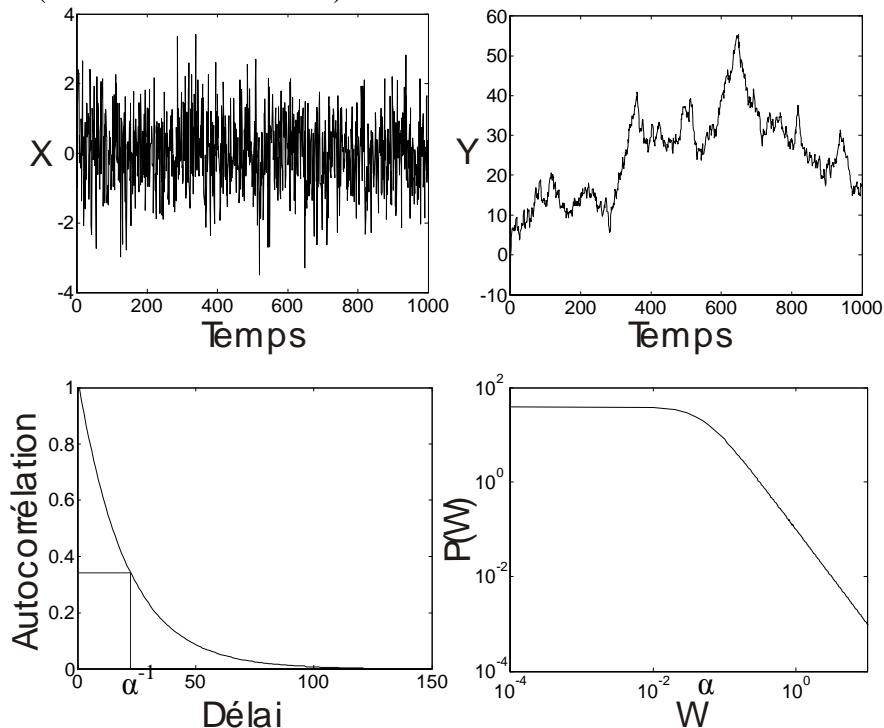
Le forçage x est considéré comme un bruit blanc par rapport à y , i.e. il est considéré sans mémoire ("memoryless") aux échelles de temps de la dynamique y . En termes de corrélation et de constante de temps, ceci revient à $\tau_x \ll \tau_y$ et $c_x(\tau_y) \approx 0$. La fonction d'autocorrélation s'obtient alors:

$$\frac{d}{d\tau} C_y(\tau) = \frac{d}{d\tau} \int y(t)y(t+\tau) dt = -\alpha C_y(\tau) \Leftrightarrow C_y(\tau) = c_0 \exp(-\alpha|\tau|)$$

En rappelant que le spectre de variance est la transformée de Fourier de la fonction d'autocorrélation, on a:

$$P_y(\omega) = \int c_0 \exp(-\alpha|t|) \exp(i\omega t) dt = \frac{2c_0\alpha}{\omega^2 + \alpha^2}$$

Pour $\omega \ll \alpha$, on a $P(\omega) \sim 1$, i.e. y se structure comme un bruit blanc aux grandes échelles de temps. Pour $\omega \gg \alpha$, on a $P(\omega) \sim \omega^{-2}$, ce qui correspond à la loi puissance du spectre d'un bruit brun (i.e. une marche aléatoire).



Kaitala et al. (1997a), et Bjornstad et al. (2004) suivent une démarche analogue dans l'analyse fréquentielle de différents modèles, et montrent que la simple connaissance du Jacobien de la dynamique (i.e. ses propriétés de stabilité) suffit à préciser la forme générale du spectre de la population:

$$x(t+1) = Jx(t) + Gu(t) \Leftrightarrow P_x(\omega) = (I - e^{i2\pi\omega} J)^{-1} G P_u(\omega),$$

où I est la matrice identité, J et G les Jacobiens du système en x et u , et P_u le spectre de fréquence du forçage u .

La séparation des échelles de temps caractéristiques permet donc d'appréhender l'effet d'un système biologique comme un filtre altérant les propriétés du forçage environnemental (Bjornstad et al. 2004). L'exemple ci-dessus correspond à un filtre dit passe-bas, de fréquence caractéristique α , atténuant les hautes fréquences pour ne conserver que les basses fréquences. La **fonction de transfert**, analogue au cospectre du terme de forçage et de la variable de sortie, synthétise ainsi la réponse fréquentielle du système (physique ou biologique) à un forçage (déterministe ou stochastique). Elle délivre en particulier de précieux renseignements sur les origines endogène et exogène du signal mesuré, et sur les interactions entre le système et le forçage auquel il est soumis. La fonction de transfert d'un système peut ainsi être obtenue de façon empirique en lui appliquant un signal stochastique exactement blanc (*i.e.* $P(\omega)=1$), puis en calculant le cospectre des variables U et Y .

Formulation alternative de la fonction de transfert

La transformée de Laplace

$$F(s)=L[f(t)]=\int_0^{\infty} f(t)e^{-st} dt$$

permet de transformer un système d'équations différentielles d'ordre N en un polynôme d'ordre N :

$$a_n y^{(n)}(t) + \dots + a_0 = 0 \Leftrightarrow a_n s^n Y(s) + \dots + a_0 = 0,$$

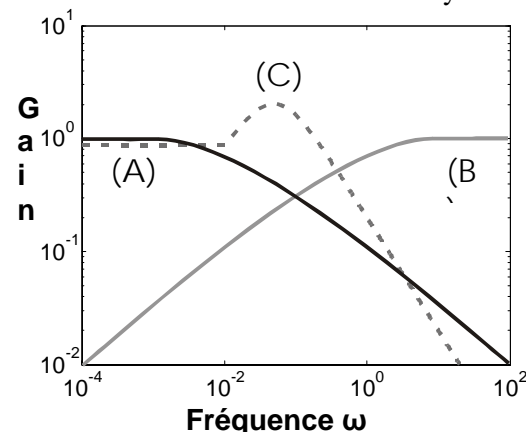
dont les racines caractérisent la forme de la réponse du système à une impulsion. Une telle opération permet d'éviter la formulation analytique (laborieuse) de la fonction de gain d'un filtre (voir encadré précédent). Un système d'équations différentielles reliant un forçage u à une sortie y s'écrit donc:

$$\begin{aligned} a_n y^{(n)}(t) + a_{n-1} y^{(n-1)}(t) + \dots + a_0 &= b_m u^{(m)}(t) + b_{m-1} u^{(m-1)}(t) + \dots + b_0 \\ a_n s^n Y(s) + a_{n-1} s^{n-1} Y(s) + \dots + a_0 &= b_m s^m U(s) + b_{m-1} s^{m-1} U(s) + \dots + b_0 \\ [a_n s^n + a_{n-1} s^{n-1} + \dots + a_0] Y(s) &= [b_m s^m + b_{m-1} s^{m-1} + \dots + b_0] U(s) \end{aligned}$$

Soit:

$$\frac{Y(s)}{U(s)} = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_0}$$

De l'ordre et des racines des polynômes du numérateur et du dénominateur dépend la forme de la fonction de transfert du système:



Fonctions de transfert (ou spectre caractéristique) de différents filtres

- (A) de type passe-bas,
- (B) de type passe-haut,
- (C) passe-bas résonant.

La formulation rigoureuse du modèle est nécessaire pour procéder à une telle analyse. Les systèmes dynamiques, comme les formulations matricielles dites "de Leslie" (Buckland et al. 2004) sont particulièrement propices à ce genre d'étude. Bjornstad et al. (1997, 2004), Kaitala et al. (1997) et Ripa et al. (1998) ont ainsi employé des techniques de linéarisation pour formuler analytiquement la fonction de transfert de différents modèle de population, de façon analogue à Milotti (1995). Le résultat majeur commun à ces études est que la structure d'une population (incluant les traits d'histoire de vie de l'espèce et les interactions entre classes d'âge) est à même de filtrer le forçage environnemental (stochastique) de façon non triviale. En particulier, un forçage complexe (en terme de composition fréquentielle) suffit à générer des cycles de grande longueur d'onde. Si les interactions entre cohortes s'apparent à un "résonateur" écologique (Bjornstad et al. 2004), celles-ci se comportent statistiquement comme un filtre passe-bas, *i.e.* de gain non-nul sur la plage des basses fréquences. L'interaction avec les phénomènes de régulation eux-même (densité-dépendants) complique le tableau et conduit potentiellement à des gains supérieurs à 1 pour certaines plages de fréquence, et conduit au phénomène de **résonance** *sensu stricto*. En d'autres termes, certaines plages fréquentielles cachées ou secondaires sont extraites du signal forçant (à ne pas confondre avec la résonance stochastique, qui concerne l'amplitude du bruit d'entrée et non son spectre).

Quelle peut être la contribution de ces travaux à la présente réflexion? Ils établissent en fait la possibilité d'une origine endogène pour les fluctuations à long terme chez le thon rouge. L'étude théorique menée par Bjornstad et al. (2004) indique que l'interaction de la structure en âge de sa population (15 à 20 classes d'âge) et d'un forçage stochastique blanc suffirait à générer des pseudo-cycles de biomasse. Ce phénomène, dénommé "résonance de cohorte" par ces auteurs, peut donc potentiellement expliquer la basse fréquence observée. C'est l'hypothèse **endogène**. L'alternative, ou hypothèse **exogène**, impliquerait un terme de forçage, potentiellement lié au climat océanique, mais de spectre analogue (*i.e.* de grande variance sur la période séculaire). En l'absence de modèle satisfaisant basé sur la biologie de *Thunnus thynnus*, il apparaît difficile de favoriser une de ces hypothèses: on doit donc considérer conjointement la possibilité d'une origine endogène et exogène à ces pseudo-cycles. Nous sommes là dans un cas de test multiple (Hilborn and Mangel 1997) qui, pour être mené à bien, doit reposer sur la construction d'un critère objectif commun, *i.e.* les deux hypothèses doivent être comparables.

Cycles exogènes? Si la population de thon rouge présente les caractéristiques générales d'un filtre passe-bas (Fromentin 2001, Fromentin and Fonteneau 2001), on s'attend à ce que les basses fréquences du forçage soit peu ou pas déformées par ce filtre, à un décalage de phase près. Les indices climatiques candidats, sous couvert d'une telle hypothèse, auraient donc un spectre de variance présentant une structure séculaire. Or une certaine tendance en écologie quantitative

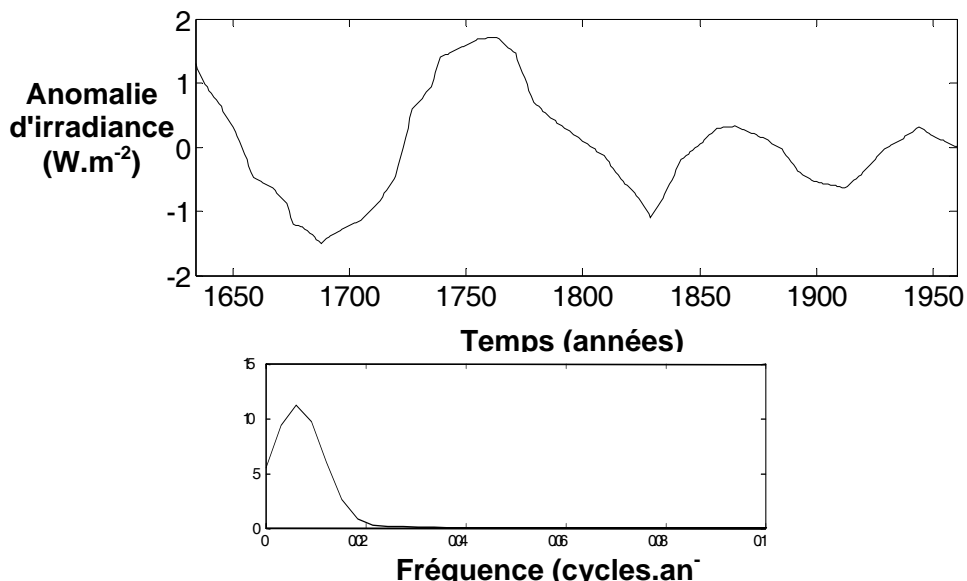
est d'assimiler les variations du climat aux fluctuations de variables environnementales locales (Stenseth et al. 2003). La rareté de séries d'indices climatiques globaux en est la principale cause. De tels indices, comme l'Indice de Durée du Jour, les Indices d'Oscillation Nord-Atlantique (Hurrell et al. 2001) ou la température atmosphérique de l'hémisphère nord (Parker et al. 1992, Jones 1994) ont pu néanmoins être inclus dans les travaux précédents (Ravier and Fromentin 2004). Des relations systématiques et fortes avec la température ont ainsi pu être mises en évidence, conduisant à l'hypothèse d'un changement de route migratoire (*i.e.* un effet à grande échelle du climat, donc exogène).

Paléoclimatologie et irradiance solaire

La reconstruction de l'histoire du climat océanique fait intervenir conjointement les outils de modélisation (*e.g.* modèles de circulation générale, ou de budget énergétique) (Cubasch et al. 1997, Crowley 2000) et des séries d'indices climatiques indirects (basés *e.g.* sur des carottages).

Un aspect majeur du climat océanique est l'évolution de la **hauteur de mer** (van de Plassche et al. 1998): elle témoigne de variations de la pression atmosphérique (l'effet "baromètre inverse"), de variations d'intensité du Gulf-Stream (Greatbach et al. 1991), de fluctuations du stress éolien, de variations stériques... Ces évolutions ont des origines internes, dues à des mécanismes mus par la densité, ou externes, liés à la variabilité du forçage **solaire**. Un grand nombre de travaux pointe ainsi l'importance de ce dernier forçage sur le climat (Haigh 1996, Reid 1997, Lean and Rind 1998), et le couple océan-atmosphère, en particulier la température des couches marines superficielles (Barnett 1989, White et al. 1997). L'hypothèse mécaniste en vigueur est celle d'un effet double de l'irradiation solaire sur la stratification de la colonne d'eau, menant à des effets immédiats (variations à haute fréquence de la température superficielle) et en grande partie à des effets retardés (de 20 à 100 ans), par l'intermédiaire de l'apparition et de l'advection d'anomalies durables de salinité (van der Schrier et al. 2002, van de Plassche et al. 2003).

Un grand nombre de reconstructions de séries d'irradiance existe dans la littérature (Hoyt and Schatten 1993, Reid 1997, Bard et al. 2000), et ont été corrélées *e.g.* à des séries historiques de température océanique de surface (Lohmann et al. 2004, Jiang et al. 2005). La série de Bard et al. (2000), basée sur l'estimation des anomalies de ^{14}C et ^{10}Be cosmogènes, est employée ici, et possède une remarquable structure quasi-périodique à 90-140 ans.



Une série reconstruite (Bard et al., 2000) décrivant les évolutions de l'irradiance solaire fournit une nouvelle variable candidate pour expliquer les fluctuations à long terme de *Thunnus thynnus*. Sa structure fréquentielle (*i.e.* trois pseudo-cycles de période 90-140 ans) apparaît remarquablement analogue à celle des séries synthétiques issues des méta-analyses précédentes. Des récents travaux de modélisation (van der Schrier et al. 2002, van de Plassche et al. 2003) indiquent un effet important du forçage solaire sur la circulation thermo-haline (*i.e.* par la production d'anomalies durables de salinité et un impact sur la dérive nord-atlantique et la température en Atlantique Nord-Est). Ceci suggérerait un lien possible (bien que très indirect) avec l'abondance de la population de thon rouge, ou l'intensité du flux migratoire touchant les zones de ponte en Méditerranée.

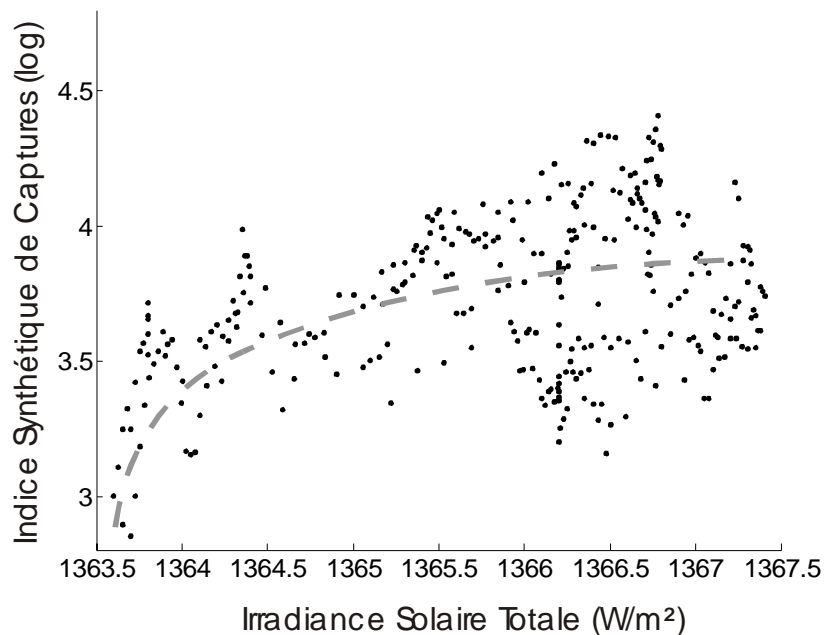


Figure 4-10. Régression de la série d'Irradiance Solaire Totale reconstruite par Bard et al. (2000) contre la série synthétique de captures estimées par EOFs, avec un délai de +20 ans.

Bien que séduisant par sa cyclicité (fournissant des "replica") et sa période apparemment adéquate, un tel terme de forçage dans un modèle climat-population soulève deux problèmes majeurs:

[1] La proposition, l'implémentation et le test d'une hypothèse basée sur des mécanismes géophysiques et biologiques, si elle est *a priori* possible malgré "l'éloignement écologique" de ces variables (l'une étant cosmogène, l'autre une série de captures liées à l'activité de pêche), demandera un effort certain pour un résultat relativement trivial. En effet, le léger décalage de phase observé entre les deux séries (environ 20 ans) serait la caractéristique majeure à reproduire: l'essentiel du signal basse fréquence devrait lui être propagé linéairement, et conservé en grande partie par le modèle explicatif. Ceci apparaît être beaucoup de travail pour ... construire un filtre linéaire. Il semble donc plus raisonnable, à ce stade du raisonnement, de se restreindre au seul exercice de formulation de

l'hypothèse (*i.e.* du recensement des mécanismes pouvant donner lieu à un tel délai).

[2] Plus important encore, le test d'une corrélation entre les deux variables devient particulièrement difficile car leur relation intervient à très basse fréquence: les deux séries étant chacune fondamentalement autocorrélée ("rouges" voire "brune" pour l'irradiance), il est nécessaire de fortement corriger le nombre effectif de degrés de liberté intervenant dans la régression (Bartlett 1946), ce qui supprime la quasi-intégralité des relations détectées à ces échelles (Ravier and Fromentin 2004). La méthode alternative serait de préserver le spectre fréquentiel de la donnée et effectuer une randomisation de type Monte-Carlo (par réarrangement des phases dans le domaine de Fourier): les séries étant quasi-cyclique, ceci revient à tester l'existence d'un décalage de phase maximisant la corrélation entre séries, qui existera presque toujours! En d'autres termes, régresser deux séries quasi-périodiques et de même période équivaut à ne travailler que sur la caractéristique qui les sépare, *i.e.* leur décalage de phase, dont la significativité n'est plus testable en l'absence d'*a priori* solide sur les constantes de temps des mécanismes sous-jacents.

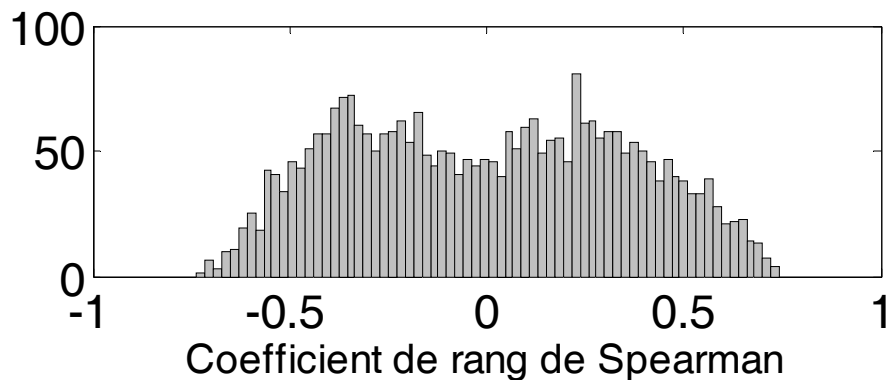


Figure 4-11. Distribution empirique du coefficient de Spearman reliant la série synthétique de captures extraite par EOF à la série d'irradiance solaire (Bard et al. 2000). Un ensemble de 3000 homologues statistiques de type 3 a été généré par bootstrap des phases dans le domaine de Fourier. La bimodalité de l'histogramme en 0.4 et -0.4 rend la statistique de test originale non significative.

Cette constatation frustrante s'applique à toute tentative de mise en relation du climat et populations marines sur de grandes échelles de temps (donc de changement climatique). On illustre ici une limite des méthodes statistiques classiques⁴ et des connaissances nécessaires au débrouillage des origines endogènes et exogènes des fluctuations écologiques⁵.

[3] Enfin, la sélection de l'une ou l'autre des hypothèses initiales (endogène contre exogène) reste difficile, malgré la grande similitude

⁴ "Red noise is hard to beat" (Allen et Robertson 1996)

⁵ qualifiée ainsi par Jonzen et al. (2002): "the irreducible uncertainty of the demography-environment interaction in ecology"

spectrale des deux séries. En effet, la formulation de la première hypothèse (endogène) implique un modèle passe-bande, voire résonant, mais, par définition, ne dépend pas de la structure exacte du forçage environnemental (il suffit qu'il soit suffisamment blanc), alors que la seconde dépend de l'existence d'une structure quasi-séculaire dans ce terme de forçage. Ce test d'hypothèse multiple ne peut pas être résolu avec les critères Bayésiens actuels de sélection (Akaike 1973, Hilborn and Mangel 1997), puisqu'ils reposent seulement sur la structure des modèles à comparer, et non la structure des forçages auxquels ils sont soumis. Un tel critère combinant ces deux aspects reste à inventer, afin de résoudre le problème de l'origine des fluctuations écologiques (Jonzén et al. 2002).

4.4. Discussion sur l'analyse des séries temporelles

Un nouveau regard a été posé dans ce chapitre sur les fluctuations d'abondance du thon rouge: la problématique a ainsi pris la forme d'un questionnement sur le degré de déterminisme des séries, leur structure fréquentielle, et l'origine endogène ou exogène des patterns identifiés. On a en particulier montré le caractère non-aléatoire des séries de capture, l'existence d'un pattern d'autocorrélation non trivial à haute fréquence, et la correspondance remarquable avec une variable à basse fréquence (reconstruite) intervenant dans le forçage de la circulation océanique.

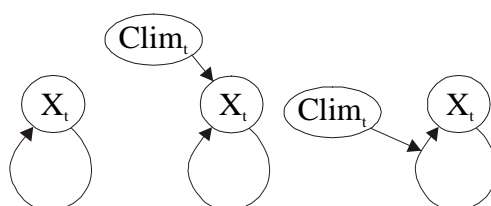


Figure 4-12. Les trois modes élémentaires de variabilité pour une population unique soumise à une densité-dépendance et un effet du climat. D'après Stenseth et al. (2002).

Densité-dépendance. Une densité-dépendance à la fois directe et retardée (*i.e.* de lag 1 et 2) a donc été démontrée. En l'absence de connaissances précises sur les chaînes trophiques associées au thon rouge, il est peu probable que l'approche suggérée par Bjørnstad et al. (1995) et appliquée avec succès dans certains travaux (Bjørnstad et al. 1995, Bjørnstad et al. 1996, Saitoh et al. 1997, Stenseth et al. 1999) permette ici une explication. Le caractère hautement opportuniste de ce prédateur, sa grande espérance de vie et sa grande mobilité tendra en effet à brouiller toute relation trophique forte avec une ou des espèces précises de proies (l'effet résultant s'apparentant à un lissage spatial et temporel). Une hypothèse alternative, encore peu évoquée dans la littérature traitant de la densité-dépendance statistique, est que l'ordre 2 du système autorégressif ne soit pas lié à des interactions entre classes d'âge ou entre espèces, mais à la somme des effets de mécanismes reproductifs. Une **densité-dépendance de reproduction** pourrait ainsi être due à une incapacité de la part des reproducteurs à assurer une

maturation et une émission d'œufs tous les ans. Le présent résultat (*i.e.* un modèle d'ordre 2, donc un pattern pseudo-cyclique de période 3) indiquerait même une reproduction intervenant tous les 3 ans en moyenne (en fait une migration liée à la reproduction, donc un flux d'adultes aux gonades matures). Ceci permettrait d'expliquer de façon relativement confortable certains résultats intrigants fournis par les marques archives: des individus potentiellement matures ont été repérés dans des zones autres que la Méditerranée ou le golfe du Mexique en pleine période de reproduction (*i.e.* de mai à août), conduisant certains auteurs à postuler l'existence de zones de ponte en plein océan (Lutcavage et al. 1999). L'explication n'impliquerait donc pas l'écologie de *Thunnus thynnus* (*i.e.* sa stratégie écotogéographique pour l'émission des œufs), mais serait d'ordre **énergétique**: il lui serait ainsi difficile de compléter chaque année le budget énergétique nécessaire à la fois à son métabolisme basal, son activité de nage et de chasse et la maturation des gonades. La longévité et la grande distribution spatiale des adultes confèrent naturellement une grande hétérogénéité aux taux de croissance observés (Fromentin 2003). Le facteur de condition associée à la reproduction pourrait en être grandement affecté.

Un bémol doit toutefois être formulé: si le mécanisme conduisant à cette densité-dépendance est effectivement individuel (et non émergent), on s'attendrait à ce qu'un tel signal soit brouillé à l'échelle de la population (*i.e.* en l'absence d'un facteur synchronisant, la somme des nombreuses cohortes venant se reproduire en Méditerranée devrait conduire à un effet distribué, potentiellement structuré en basse fréquence mais non en haute fréquence). Le comportement migratoire de l'espèce étant une caractéristique majeure, il est probable qu'un effet climatique local, variant à haute fréquence, puisse avoir un effet synchronisant sur la fréquence des événements reproductifs (eux-même limités par l'interaction physiologie/environnement) en altérant le système de navigation pélagique des thonidés (Walker 1984, Carey 1990, Horvath and Varju 1995). L'importance du comportement n'est pas non plus à négliger: si le déclenchement et l'étendue de la migration chez *Thunnus thynnus* possède une dimension d'apprentissage, même minime (*i.e.* un transfert de connaissances entre individus ayant ou non migré, par l'intermédiaire du banc), il est possible d'imaginer que l'unité du banc suffise à induire un effet stabilisant, synchronisant les migrations de reproduction à l'échelle d'une sous-population, voire plus. Cet effet stabilisant, peu éloigné de la "school-trap hypothesis" proposée par Bakun (1999) pour les petits pélagiques, est en accord avec les propositions de Steele (1985, 1989), *i.e.* la structure particulière des écosystèmes marins est apte à générer des patterns forts et de grande variance sans pour cela nécessiter des causes aussi nettes et évidentes qu'en terrestre. En d'autres termes, la grande non-linéarité du biotope marin suffit à générer des effets disproportionnés aux causes. Cette hypothèse (une asymétrie dans la reproduction combinée à un facteur synchronisant exogène ou endogène, banc ou climat) demande à être testée sur le terrain.

Enfin, une remarque méthodologique s'impose sur l'interprétation d'une densité-dépendance statistique avec retards ("Delayed Density-Dependence", ou DDD). Un tel phénomène, lorsqu'il est détecté, est souvent la signature d'interactions trophiques remarquables de type prédateur/proie, ou de boucles de rétroaction entre juvéniles et adultes (Royama 1992, Berryman and Turchin 1997, Fromentin et al. 2001). Toutefois, Jiang et Shao (2004) ont très récemment montré comment la structure fréquentielle de l'environnement (*i.e.* un signal rouge, autocorrélé) suffisait à induire des patterns fortuits de densité-dépendance retardée, lorsque seule une densité-dépendance directe était formulée pour les jeunes classes d'âge. En d'autres termes, c'est encore l'interaction entre structure fréquentielle du forçage et structure des interactions intrinsèques qui brouille le problème en dynamique des populations (Williams and Liebhold 1995). Si le signal qui a été mis en évidence par différentes méthodes d'autocorrélation et de méta-analyse est effectivement fort et très significatif, son interprétation reste difficile et subordonnée à de plus amples informations.

Basse fréquence. L'hypothèse originelle posée par Ravier et Fromentin (2004), postulant un effet du climat sur la dynamique du thon rouge, est à la fois renforcée et nuancée. Un effet direct et visible du climat sur le recrutement de *Thunnus thynnus* reste ainsi peu probable, au vu de la grande mobilité des adultes et de leur faculté à exploiter les mers adjacentes de l'Atlantique Nord (littéralement des océans miniatures, offrant un grand choix d'options thermiques au sein de zones restreintes). L'effet sur la distribution océanique des adultes (et une interaction avec les migrations) reste l'option principale, en particulier avec l'inclusion de la série reconstruite d'irradiance solaire montrant une cyclicité remarquablement analogue aux fluctuations de la population.

Une hypothèse complémentaire, plus classique, ferait intervenir un effet Moran (donc un effet distribué, à grande échelle, sur la survie des adultes) interférant avec les migrations de reproduction en direction de la Méditerranée. La capacité de charge globale du milieu serait ainsi affectée de façon cyclique, et contraindrait simplement le nombre d'adultes survivants chaque année. En l'absence de séries concomitantes de production biologique en Atlantique nord sur les 300 dernières années, ou de modèle simple liant irradiance et productivité océanique, cette hypothèse reste à tester.

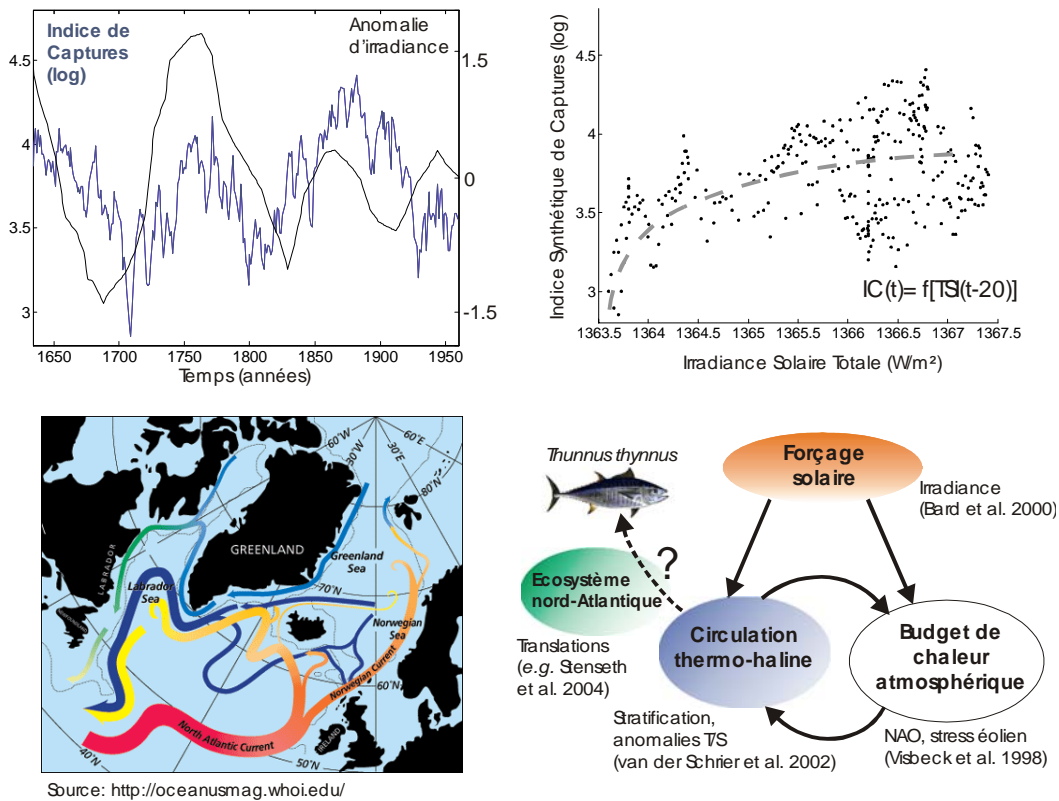


Figure 4-13. Un mécanisme "bottom-up" liant le forçage solaire aux fluctuations de la population de *Thunnus thynnus*?

La mise en corrélation de séries temporelles de spectre coloré (*i.e.* dominé par les basses fréquences) reste un exercice statistique difficile. Phase, période et amplitude peuvent se combiner pour rendre les approches classiques inutiles. En particulier, deux séries dont les phases sont cohérentes peuvent apparaître décorréliées si leurs amplitudes respectives varient librement. Pour tester plus avant un effet possible du climat, l'analyse des phases (Cazelles and Stone 2003) apparaît être une étape importante dans la suite. L'effet du faible nombre de *replica* (*i.e.* trois pseudo-cycles sur la période 1634-1960) sur la robustesse de l'analyse doit en particulier être étudié.

Nouvelles pistes. La prise en compte de la structure fréquentielle des forçages environnementaux est encore peu répandue dans les travaux d'écologie marine (McCann et al. 2003). Inspirée des travaux récents sur la "couleur du bruit environnemental"⁶ (Kaitala et al. 1997, Petchey 2000, Ranta et al. 2000, Vasseur and Yodzis 2004), une étude a été menée sur la structure fréquentielle des mesures de température de surface issues de la base COADS (<http://www.cdc.noaa.gov>). Les séries ont été extraites sur la période 1900-2002 puis agrégées sur 8 régions biogéographiques recouvrant la distribution spatiale des adultes de *Thunnus thynnus* (Longhurst 1998). Si

⁶ "The colour of environmental noise" résume l'incapacité des modèles classiques à générer des fluctuations à basses fréquences telles qu'observées dans les séries écologiques (*i.e.* de spectre rouge).

la température de surface reste un indice très synthétique de la variabilité océanique, elle reste un facteur majeur dans le développement et la croissance des larves de thonidés, pour lesquelles des valeurs de plus de 24 °C sont nécessaires (Schaefer 2001). Le **manuscrit 6** résume cette étude et ses résultats: on y montre en particulier les différences de ratio entre hautes et basses fréquences pour les zones océaniques et pour les mers adjacentes. La Méditerranée en particulier montre une pente spectrale moins forte (plus proche d'un bruit rose, ou "pink noise", Morales 1999) que les autres régions: sa qualité d'océan miniature, à la réponse rapide aux forçages atmosphériques, offrirait un vaste choix d'options (en termes reproductifs) dans des fenêtres spatio-temporelles limitées. La valeur d'adaptation évolutive de la "stratégie" du thon rouge venant se reproduire aux mois d'été dans cette zone peut en particulier être discutée. Si ce travail reste préliminaire, il constitue une voie intéressante pour la recherche et l'étude des effets climatiques sur la variabilité des populations marines.

Méthodologie de l'enquête. La démarche employée dans ce chapitre innove par la remise en cause de deux paradigmes majeurs, qui postulent que: (i) des mécanismes écologiques ou géophysiques distincts ont nécessairement des signatures fréquentielles distinctes, et (ii) la modélisation (déterministe ou stochastique) d'un signal peut se faire pas à pas, indépendamment de sa complexité.

La théorie du signal et des systèmes dynamiques, si elle fournit un outil puissant et élégant pour l'estimation et l'identification des processus cachés, est fondamentalement dépendante d'un *a priori* sur le système sous-jacent (cet *a priori* prenant la forme d'un modèle). Il est donc hasardeux d'employer ces techniques à des fins exploratoires si l'on ne se place pas dans un cadre de test d'hypothèses multiples. Par exemple, la distinction entre dynamique et bruit d'observation est implicitement effectuée en terme fréquentiel: le bruit d'observation est réduit à un bruit blanc (par commodité mathématique, Donoho 1992) et est donc caractérisé par un spectre fréquentiel uniforme. Le danger est grand, dans le cas de méthodes d'ajustement sophistiquées (e.g. échantillonneur de Gibbs), de voir attribuée la composante haute fréquence de la dynamique réelle à ce processus d'observation. Or, la vérification des résidus, étape essentielle à la modélisation, porte souvent sur l'étude de leur autocorrélation (*i.e.* l'absence de tendance ou de pattern global), un concept inapte, on l'a vu, à rendre compte de structures particulières dans les hautes fréquences. Le rapport signal-bruit est donc une notion qui doit être considérée sur l'ensemble du spectre fréquentiel, et non être ramenée à l'approximation "haute fréquence=bruit non informatif, basse fréquence=dynamique d'intérêt".

Pour pallier à ces problèmes, l'analyse exploratoire des séries de capture a été approfondie, en appliquant certaines techniques employées dans l'analyse des dynamiques chaotiques. Une **modélisation exploratoire** a ainsi été menée. Un retour aux définitions premières de la linéarité et de l'aléatoire permet ainsi d'apprécier les propriétés

fondamentales de ces séries. On teste ainsi, par la méthode des homologues statistiques, ou "surrogates", la possibilité d'un déterminisme et de relations non-linéaires sans définir *a priori* le modèle exact. Ce test d'hypothèse **composite** est rendu possible par la récente mise au point des méthodes de simulation spectrale, conservant la structure fréquentielle du signal mais annihilant toute structure déterministe ou non-linéaire (Casdagli 1989, Theiler et al. 1992, Breakspear et al. 2003, Angelini et al. 2004). Nos séries de captures étant remarquablement longues, une méthode exotique basée sur la construction d'un espace des phases empirique et la quantification des patterns **récurrents** est proposée. Cette méthode, si elle apparaît relativement puissante lorsque la donnée est non-limitante, permet d'illustrer un pré-requis capital en écologie quantitative: il n'y a de déterminisme ou d'expérimentation en séries temporelles que sous réserve de répétition. Des états multiples, s'ils sont chacun peu observés, fournissent alors une information mineure sur la dynamique: si le système est constamment sujet à une dérive (e.g. dans le cadre du changement climatique), l'identification des processus ne peut se reposer sur la recherche empirique de son déterminisme, mais sur sa formulation préalable et son test. La méthode des **réurrences** reste une voie prometteuse pour la modélisation exploratoire et non-paramétrique de systèmes complexes. Il est ainsi possible d'envisager la détection d'attracteurs (Zbilut et al. 2002), la mise en valeur de dérives (Atay and Altintas 1999) ou tout simplement l'extraction de patterns pseudo-cycliques, à l'instar des ondelettes (Torrence and Compo 1998). Lorsque la donnée est limitante, les méthodes d'augmentation et d'interpolation (Eckmann and Ruelle 1985) permettent l'extraction de dynamiques non-triviales sous réserve, bien sûr, de leur existence réelle (Boudjema and Cazelles 2001). Ces méthodes de modélisation non-linéaires sont encore peu développées en écologie (Pascual and Ellner 2000), et restreintes souvent e.g. au calcul du coefficient de prédiction (Turchin 2000, Turchin and Ellner 2000). Elles restent toutefois promises à un avenir intéressant: en effet l'ajustement d'un modèle dans l'espace d'état (empirique/reconstruit grâce au théorème de Takens) est privilégié par rapport aux ajustements dans le temps (e.g. basé sur une fonction d'autocorrélation, une dynamique temporelle...). On profite ainsi du maximum d'information qu'il est possible d'obtenir de la dynamique observée, *i.e.* l'existence et la valeur d'états passés (proches dans l'espace d'état mais éloignés dans le temps) sont ainsi exploitées. L'emploi d'une distance euclidienne pour pondérer la contribution de chaque état proche est alors tout à fait envisageable, puisque celle-ci est homogène en terme de dimensions (*i.e.* les états X_i sont formés comme une agrégation successive d'une même variable).

Enfin, le test d'hypothèses multiples, basé sur une série de **problèmes inverses** apparaît comme une alternative cohérente à la modélisation exploratoire (Hilborn and Mangel 1997). Toutefois, on l'a vu, la formulation des différentes hypothèses devient problématique: une ou plusieurs hypothèses postulent ainsi une origine **endogène** aux fluctuations, donc peu dépendante du terme de forçage (Kendall et al. 1999, Lundberg et al. 2000) contre une origine **exogène**, dépendante des

caractéristiques de ce forçage (Ripa et al. 1998, Ranta et al. 2000). Le critère de sélection des modèles correspondants étant uniquement basé sur ces mêmes modèles et non leurs variables de contrôle (Akaike 1973, Hilborn and Mangel 1997, Burnham and Anderson 1998), le test objectif n'est plus possible et conduit à nuancer les conclusions finales de Jonzén et al. (2002) sur leur emploi. Une voie de recherche intéressante s'ouvre ici, impliquant peut-être des méthodes de simulation exacte ou la formulation de nouveaux critères de sélection.

5. Synthèse et perspectives

Nous ne raisonnons que sur des modèles

Paul Valéry

5.1. Synthèse sur l'objet de recherche, *Thunnus thynnus*

[1] Les fluctuations dans le temps et l'espace des populations marines résultent de phénomènes internes au système (*i.e.* les processus biotiques des écosystèmes) et externes (*i.e.* dus au forçage climatique, la pression de l'activité humaine par pêche ou pollution, ou des effets combinés pêche-climat), ou d'une interaction entre ces effets.

Les mécanismes supposés de ces fluctuations, représentés dans des diagrammes temps-espace, ont des ordres de grandeurs variant quasi-linéairement: leurs échelles spatio-temporelles varient conjointement et continûment. Les espèces de niveau trophique supérieur comme le thon rouge y occupent souvent une place restreinte, à l'extrémité de ces spectres. Pourtant le diagramme de variabilité océanique montre que de telles espèces, très mobiles, peuvent être potentiellement affectées par une grande variété de mécanismes et d'échelles.

Les modalités des interactions thon rouge/environnement doivent être donc abordées à ces **différentes échelles spatiales et temporelles**: le présent travail en aborde trois (*i.e.* locale, régionale et globale). Plusieurs types particuliers de données ont été employés pour cela (*i.e.* des comptages par survol aérien, des mesures de radiométrie satellitale, des enregistrements de marques archives, des sorties d'un modèle de circulation océanique, des données historiques de captures par madragues, et une série temporelle reconstruite d'une variable climatique). Les spécificités de chaque étude ont conduit à développer ou adapter des méthodes d'analyse relativement nouvelles en écologie marine: leur intérêt était de permettre à la fois l'analyse exploratoire (*i.e.* l'extraction d'information non spécifiée *a priori*) et le test d'hypothèse (*i.e.* l'acceptation ou le rejet d'un modèle nul).

[2] A l'échelle locale, la relation bancs/fronts, si souvent évoquée dans la littérature halieutique, a été examinée. Les fronts sont des objets de recherche relativement complexes: si certaines structures frontales stables sont assez bien connues en océanographie physique (*e.g.* les fronts de panaches, les fronts planétaires, les fronts de marée), la majorité des fronts océaniques, fugaces et intermittents, reste peu étudiée ou même cartographiée. Les observations et les modèles indiquent pourtant qu'ils peuvent avoir un impact non négligeable sur les échanges océan-atmosphère, les mouvements d'eau verticaux et les autres processus majeurs affectant le biotope (*e.g.* enrichissement-concentration-rétention). Une brève introduction à la physique des fronts a été effectuée au chapitre 2: la grande diversité de leurs mécanismes potentiels limite leur **prédictibilité** par des modèles de circulation, ainsi que la compréhension

des relations exactes liant les fronts et les communautés planctoniques et nectoniques. Il reste avéré que la grande hétérogénéité spatiale des zones frontales structure l'habitat pélagique et sa production biologique: leur cartographie systématique et à haute résolution par des moyens satellites est donc indiquée pour étudier et expliquer la distribution spatiale de nombreuses espèces pélagiques. La mise au point **d'algorithmes statistiques** de détection des fronts apparaît aussi nécessaire pour faciliter et accélérer de telles études. Enfin, les mesures disponibles étant spatialisées, on doit employer une méthode appropriée de test d'hypothèse, *i.e.* permettant de quantifier l'écart du pattern observé à un pattern de distribution aléatoire.

On a testé ainsi l'existence d'une relation entre la distribution spatiale des juvéniles de thon rouge et les structures de surface dans le Golfe du Lion. Cette zone présentant un fort signal océanographique (interactions entre un courant permanent, un panache fluvial et un régime de vents intense), on s'attendait à détecter ces structures par des mesures indirectes de télédétection. Les plus hautes résolutions spatiales disponibles (*i.e.* 1.1 km) ont été employées, en conjonction avec les positions GPS des bancs repérés lors de la campagne STROMBOLI. La modélisation de ces observations a été effectuée grâce à la géostatistique et au formalisme des processus ponctuels. Le calcul de divers indices de dispersion et de variance ont mis en évidence des différences significatives de structure spatiale par rapport à une distribution nulle (*i.e.* aléatoire, ou de Poisson). La prise en compte d'un effet de l'environnement et des zones frontales sur la variable mesurée a ensuite permis d'expliquer en partie les patterns observés. Si ces patterns étaient globalement compatibles avec un effet de l'environnement sur des échelles de 10 à 40 km, les échelles inférieures (<10 km) montraient un surplus d'agrégation, *i.e.* l'existence d'interactions entre bancs ou avec les densités locales de proies. Les échelles supérieures (>40 km) montraient un surplus de dispersion (ou un déficit d'agrégation) par rapport au modèle nul, posant le problème de la dynamique des individus à l'échelle de la zone.

Si effectivement la relation prédateurs/fronts est encore une fois démontrée, on montre ici le besoin d'une approche dynamique, et non plus statique au problème. Le cadre de cette étude dépasse enfin la simple mise en corrélation biologie/environnement: les zones frontales sont en effet des zones privilégiées d'exploration pour les pêcheries pélagiques. L'emploi de l'imagerie satellitale à des fins d'aide à la pêche est maintenant généralisé: les moyens de transmission actuels mettent à disposition des capitaines de pêche des séries quotidiennes d'images à haute résolution de la dynamique océanique. En d'autres termes, les structures de surface et leurs propriétés agrégatives seront responsables de fortes captures et de biais non-négligeables lors de leur analyse. Cette situation ne semble pas aussi aigüe que dans les zones tropicales (affectées en premier lieu par la pêche sous objets dérivants): si ce n'est pas (encore) un problème de pêche responsable, c'est en tout cas un biais

supplémentaires pour les analyses statistiques menées à des fins d'estimation des stocks.

[3] Cette nécessité d'une vision dynamique a été l'occasion de se pencher sur un nouvel outil d'observation, la marque archive, et surtout d'aborder des méthodes particulières issues du traitement du signal, les filtres linéaires et non-linéaires. Les marques archives ont suscité un engouement particulier en écologie marine, puisqu'elles fournissaient une solution à un problème dur d'observation: il devenait possible de suivre l'évolution d'un individu sur de longues périodes de temps, et d'enregistrer à intervalle constant des variables d'intérêt (e.g. température, profondeur). Un capteur photométrique autorise en parallèle une géolocalisation sommaire, avec une erreur de l'ordre de 1 degré. Une telle erreur de localisation est toutefois peu compatible avec un objectif d'étude de l'habitat pélagique: les récents travaux impliquant les marques archives pointent tous le besoin d'une meilleure précision de géolocalisation. L'emploi de la mesure de température ambiante semble être la solution favorisée. En effet les gradients thermiques de surface fournissent une contrainte pour améliorer la localisation des individus dans le cadre d'une estimation optimale. En spécifiant un modèle sous-jacent à la dynamique de l'individu, il devient aussi possible de tester l'existence de différents modes de déplacements dans la colonne d'eau, en relation avec des facteurs externes.

La nature non-linéaire et non-gaussienne de la variable de température et du modèle de déplacement limite le choix des outils d'estimation: les méthodes classiques, comme le filtre de Kalman simple sont en particulier inapplicables. Ses versions étendue et d'ensemble sont présentées, ainsi qu'une version généralisée basée sur une méthode de Monte-Carlo, le filtre Particulaire. Ce filtre est implémenté et testé sur un cas contrôlé: on y montre que 1) valoriser l'information sur la température permet effectivement de corriger des biais en latitude non-stationnaires ou inconnus (dérives ou détérioration du capteur), 2) des erreurs de localisation de l'ordre de la méso-échelle suffisent à générer des situations indéterminées, et 3) les performances du filtre dépendent de la bonne spécification des statistiques du champ de température employé (*i.e.* biais, erreur). Si le filtre Particulaire permet de traiter des modèles non-linéaires et des observations non-gaussiennes, sa grande souplesse nécessite donc des *a priori* informatifs, comme tout modèle Bayésien complexe.

Si les marques archives, et plus particulièrement les marques dites "pop-up", fournissent des renseignements précieux sur l'étendue des migrations à l'échelle des bassins ou sur les mouvements régionaux, elles ne semblent pas offrir la précision nécessaire pour résoudre, à échelle fine (sub-mésoéchelle), les interactions individus/habitat. La mesure apparaît ici limitante: au vu du coût de cette technologie (impliquant la balise et les jours de mer associés au marquage), il semble important de continuer l'effort d'amélioration des capteurs, voire de développer de nouveaux modes de géolocalisation (e.g. une navigation à l'estime basée sur l'enregistrement de l'azimuth et des séquences d'accélération). Les

grands programmes de marquage actuels sont aujourd'hui largement centrés sur la capture et l'équipement massif d'individus: pour maximiser leur "rentabilité scientifique", il apparaît nécessaire aujourd'hui de considérer conjointement (i) la construction de plans de marquage spatialisés et "optimaux", et (ii) la recherche constante de nouveaux moyens de mesure et de traitement de l'information.

[4] Les migrations du thon rouge couvrent de grandes distances géographiques: leurs avantages (ou leurs origines) ont déjà été évoquées en termes évolutifs, par analogie avec d'autres espèces marines migratrices. De récents travaux en écologie théorique soulignent la complexité des interactions climat/populations en séries temporelles: l'idée générale est de poser le système écologique comme un filtre, présentant une réponse particulière au forçage environnemental, à fréquence donnée. L'intérêt de cette représentation est de ne plus considérer ce terme comme un "bruit", mais comme une composante essentielle du système, de spectre complexe. Un résultat majeur concerne l'interaction entre traits d'histoire de vie d'une espèce (longévité, nombre de classes d'âge juvénile, interaction entre cohortes) et son environnement: ceci permet de proposer une explication au spectre rouge des fluctuations écologiques. Chez la population de thon rouge, le spectre des fluctuations temporelles est dominé par les basses fréquences, avec des pseudo-cycles de période 110-160 ans, et secondairement de 20-33 ans. Une hypothèse basée sur des changements de routes migratoires en relation avec la température océanique a été proposée pour expliquer les cycles de grande période.

Un nouveau pas est ici franchi dans l'étude des séries temporelles de captures: complétant l'analyse exploratoire menée dans les travaux précédents, diverses méthodes linéaires et non-linéaires sont proposées et appliquées afin d'analyser leur structure. Ces séries étant relativement longues pour un jeu de données écologiques, nous nous sommes attachés à quantifier leur déterminisme, et l'existence de patterns non-aléatoires. La méthodologie du test d'hypothèse est en particulier approfondie et raffinée. Des outils de simulation spectrale sont introduits afin de générer des homologues statistiques de la donnée observée: ceux-ci correspondent à diverses hypothèses nulles, dont l'interprétation écologique est détaillée. Les différentes séries sont ainsi testées contre ces homologues stochastiques, et montrent un surplus de patterns récurrents: l'analyse de leur structure d'autocorrélation met ensuite en évidence un phénomène de densité-dépendance directe et retardée. La reconstruction par méta-analyse de la dynamique commune aux madragues confirme et affine cette structure. Les pseudo-cycles de basse fréquence sont ensuite confrontés à un indice paléoclimatologique reconstruit, synthétisant les variations de l'irradiance solaire de 1634 à 1960. Si cette série cyclique montre une période remarquablement analogue à la série des captures, on montre que la recherche de corrélations à de telles fréquences pose un problème statistique fondamental. En effet un tel niveau d'autocorrélation met en défaut la plupart des méthodes de régression classiques (qui postule en général

des dynamiques décorréées ou blanches), et rend non significatif le test d'un décalage de phase quelconque entre séries de même période, et le test d'un signal écologique dans le cadre du changement climatique en général. Malgré ces problèmes statistiques durs, une hypothèse impliquant des variations de la circulation thermo-haline et un impact sur les communautés marines en Atlantique Nord peut être proposée. Une étude plus approfondie basée sur l'analyse des phases ou la construction d'une relation mécaniste fournirait l'exemple d'une réponse potentielle à un forçage de très basse fréquence (*i.e.* de période largement supérieure à la longévité de l'espèce). Ceci viendrait compléter l'hypothèse endogène déjà proposée dans de précédents travaux (impliquant un effet de résonance du au grand nombre de cohortes potentiellement en présence lors de la reproduction).

La possibilité d'une interaction entre traits d'histoire de vie et forçage environnemental mériterait enfin d'être étudiée plus avant: l'analyse de la base ICOADS sur l'aire de distribution de *Thunnus thynnus* illustre comment des telles analyses peuvent être menées. On montre en particulier que les fluctuations thermiques en Méditerranée sont moins variantes dans les basses fréquences que dans les zones océaniques: ce spectre plus homogène est en rapport avec la taille du bassin Méditerranéen, et de son plus faible temps de réponse au forçage atmosphérique. En d'autres termes, une mer adjacente comme la Méditerranée offre un plus grand nombre "d'options" dans une fenêtre spatio-temporelle plus réduite que les autres zones océaniques (*i.e.* son spectre est d'amplitude moindre dans les basses fréquences). Une interprétation évolutionniste peut ici être proposée avec précaution pour expliquer le décalage net entre la distribution annuelle des adultes et ces zones de reproduction particulières: l'état d'œuf ou de larve étant le stade le plus vulnérable aux conditions ambiantes chez le thon rouge, la survie des premiers âges serait favorisée dans de tels milieux, plus variés sur de faibles échelles d'espace et de temps, mais plus stables sur le long terme. Les adultes, du fait de leurs adaptations physiologiques, n'auraient pas cette contrainte et peuvent ainsi exploiter la majeure partie de l'Atlantique Nord.

Enfin, un phénomène de densité-dépendance direct et retardé est mis en évidence dans l'ensemble des séries, et confirmé dans la construction d'une série synthétique. Les processus candidats pour expliquer un tel pattern seraient d'ordre reproductif (plutôt qu'une interaction entre classes d'âge) . La possibilité d'une reproduction tous les deux ou trois ans est ainsi envisagée, essentiellement pour des raisons énergétiques (coût de la maturation des gonades). Des effets synchronisants à l'échelle de la population restent possibles (mais sans doute à court terme), comme *e.g.* un effet des conditions physiques locales (par le biais de la température sur la vitesse de maturation des gonades), ou un phénomène comportemental stabilisant (mélange au sein du banc de migrants et de non-migrants).

Les méthodes de traitement du signal ont été ici précieuses pour étudier la structure des fluctuations de la population de *Thunnus thynnus*. Un point fondamental du traitement de données est d'ailleurs soulevé, *i.e.* la distribution de l'information écologique sur l'ensemble du spectre fréquentiel, des hautes aux basses fréquences. Un *a priori* adéquat, ou au moins une analyse exploratoire poussée au préalable, est nécessaire pour ajuster tout modèle et conduire ensuite un test d'hypothèse approprié. La méthodologie du test d'hypothèse a en particulier été reconsidérée dans ces travaux: reconnaissant que l'écologie quantitative évolue vers une vision de plus en plus complexe et intégrative, il importe de développer en parallèle et sans retard les méthodes de test adéquats. Or, on l'a vu, les origines endogènes et exogènes des fluctuations écologiques sont imbriquées et difficiles voire difficile à démêler. Il convient donc de considérer à la fois le modèle ET la structure fréquentielle du forçage auquel il est soumis. Les méthodes classiques comme les critères d'information ne prennent pas en compte aujourd'hui ce problème. Dans l'attente de nouveaux critères, l'alternative proposée dans ce travail est la simulation spectrale: la construction d'homologues statistiques d'une variable de contrôle (ou du modèle lui-même!) permet d'envisager correctement le test d'hypothèse sur des systèmes complexes, caractérisés par des spectres non-triviaux (en $1/f$, voire résonants). De telles méthodes permettent de valoriser données et modèles dans un cadre de Monte-Carlo, et facilitent la recherche de dynamiques déterministes ou non-linéaires.

5.2. Perspectives

Le concept soulevé dans ces pages, **l'écosystème comme un filtre**, n'est certes pas nouveau: l'idée que les populations marines pouvaient intégrer et propager les forçages géophysiques date de près d'un siècle (Hensen 1887, Hjort 1914). La proposition qui est faite ici est d'appliquer l'ensemble de la théorie du signal et ses outils pour identifier la structure de ce filtre, et les composantes fréquentielles des forces auxquelles il est soumis. Le système réel étant peu ou pas connu, il est possible d'appliquer ces méthodes à toutes les représentations dont nous en disposons ou que nous pouvons en construire (*i.e.* données, modèles, paradigmes), afin de vérifier la validité des hypothèses qu'elles postulent. Cette proposition se décline en deux volets, l'un résolument méthodologique, l'autre orienté sur l'écologie quantitative et les problèmes d'identification. Le coût de la collecte de données en écologie est un élément fondamentalement limitant de l'enquête écologique. Une solution est l'emploi systématique des modèles pour à la fois aider à leur interprétation ou même s'y substituer pour générer des *replica in silico*. Toutefois la complexification des ces modèles et des hypothèses à tester pose un problème aigu: l'information nécessaire à leur construction, leur calibration ou leur discrimination croît continuellement et risque de dépasser les capacités d'extraction des méthodes statistiques classiques. La taille limitée des jeux de données et l'emploi encore rare des techniques de modélisation d'ensemble nécessitent de proposer une

alternative. Les auteurs de récents travaux de modélisation en physique non-linéaire (traitant de phénomènes invariants d'échelle tels que la turbulence) ont précisément défini ce besoin⁷. La difficulté de la reproduction des mesures (expérimentales ou de terrain) a donc créé ce besoin. La construction de théories et de modèles ne peuvent que partiellement le combler: un modèle véhicule rarement l'information nécessaire pour le réfuter, *i.e.* il est rarement associé à sa significativité. Les tests de sensibilité, et autres méthodes basées sur la perturbation de certains paramètres, ne peuvent être en aucun cas des tests d'hypothèse s'ils ne sont pas formellement accompagné de la formulation de l'hypothèse nulle. Par exemple, fixer $n-1$ paramètres d'un modèle à n paramètres n'a aucune réalité écologique: l'hypothèse nulle H_0 correspondante porterait sur des mécanismes stationnaires ou une dynamique parfaitement régulée voire statique. La seule hypothèse nulle intéressante est bien évidemment celle d'une nature totalement aléatoire et non-régulée pour les mécanismes sous-jacents. Cette hypothèse nulle doit donc être rejetée pour accéder à une vision déterministe du système (si elle existe). En aucun cas, l'hypothèse de départ H_0 ne peut être l'hypothèse déterministe (*i.e.* seule l'hypothèse H_1 doit contenir ce déterminisme).

[1] La construction d'homologues statistiques, par bootstrap dans un domaine autre que le temps ou l'espace, serait une solution à ce problème. Leur intérêt est de grandement améliorer l'impact des modèles dans le raisonnement scientifique, en valorisant leurs résultats dans un cadre statistique moins coûteux que la modélisation d'ensemble. Plus généralement, elle permet d'apprécier la variance de mesures expérimentales, ou de séries spatio-temporelles d'observations in-situ (l'hypothèse ergodique est alors sous-entendue). L'idée est de considérer qu'un niveau d'information suffisant est atteint pour que le spectre de variance de la données soit informatif sur la dynamique sous-jacente. Les algorithmes récents de **bootstrap** opèrent dans le domaine de Fourier, et permettent principalement la génération de signaux en $1/f$ (*e.g.* des phénomènes stochastiques opérant en cascade). Des avantages non négligeables de la transformée de Fourier sont l'orthogonalité et une grande rapidité. Toutefois cette opération suppose la stationnarité de la série et détruit toute réalité spatiale et temporelle (la dynamique) originellement présente: il convient de se tourner vers d'autres méthodes, comme la décomposition dans un espace temps-fréquence grâce à la méthode des ondelettes. Il est proposé dans le futur d'approfondir les méthodes de simulation spectrale en se basant sur les **ondelettes** et leurs propriétés **décorrélatives**. La première étape suppose le bootstrap des coefficients 1D, 2D ou 3D d'ondelettes, puis l'étude de techniques de bootstraps plus complexes, conservant *e.g.* la corrélation avec des variables externes.

Cette approche peut en premier lieu être appliquée aux séries temporelles (1D): le problème de la significativité des études sur le

⁷ "[...] a growing need for statistical methods that generate an ensemble of plausible realizations of a hierarchical process from a single a run or experiment" (Angelini et al. 2004)

changement global peut ainsi être abordé. Une tendance, une variation dans la cyclicité, ou d'autres phénomènes d'intérêt dans les séries écologiques pourraient être testés plus avant, ainsi que leur corrélation avec des indices climatiques. Ce champ de recherche étant fondamentalement limité par la qualité et la quantité des données disponibles, il importe de proposer les méthodes les plus objectives et les plus robustes.

[2] Le deuxième volet, s'il est très similaire au premier d'un point de vue méthodologique, en prend le contre-pied philosophique. On propose ici de pousser plus avant les méthodes d'identification phénoménologiques des mécanismes écologiques. Si le coût et le caractère limitant de l'observation sont criants pour de nombreuses études, il reste avéré que l'écologie marine dispose aujourd'hui d'une masse considérable de données sur la variabilité des couches de surface océanique. Le constat suivant s'impose: les modèles de circulation ou de production primaire sont vus comme suffisamment matures pour permettre des schémas d'assimilation de données poussés. En d'autres termes, la donnée est considérée dans ce cas comme subordonnée au modèle. Or un aspect essentiel des séries écologiques a été mis en évidence dans ces travaux: il s'agit du lien direct entre la taille et la résolution d'une série d'observation et son contenu informatif sur ses propres origines. L'accumulation constante des données issues e.g. de télédétection laisse imaginer que les mécanismes sous-jacents (partiellement observés) ont été sujets à un grand nombre de forçages différents (stochastiques), mais montrent aussi des états récurrents toujours plus nombreux (le déterminisme recherché). Cette croissance du contenu informatif est constante, alors que la mise au point d'un modèle, sa calibration, validation, amélioration et réfutation sont des processus toujours coûteux et plus ou moins lents. On peut donc imaginer que la progression en contenu informatif d'un jeu de données précis dépasse un jour le potentiel informatif du modèle ou des théories qui lui sont associées. Le modèle (ou paradigme), inapte à capturer d'autre variabilité que ce qui lui a été dicté, doit alors être rejeté ou remanié. Il doit donc devenir ici subordonné aux observations.

Ce courant de pensée est récent et encore ténu. La mode de l'assimilation de données, très puissante lorsqu'il s'agit d'aborder le domaine prédictif, est souvent indiquée comme l'étape prochaine dans de nombreux projets de modélisation. Pourtant elle n'est valable qu'après avoir répondu à la simple question: "la théorie disponible est-elle un objet plus informatif que le jeu d'observations disponible?". Dans le cas d'une réponse affirmative, l'assimilation est toute indiquée. Que faire dans le cas contraire? Notre incapacité à résoudre de nombreux phénomènes physiques et écologiques montre que cette question n'est pas anodine.

Une approche statistique des modèles et des observations peut être ici proposée, faisant usage de l'ensemble des outils du traitement du signal et des systèmes dynamiques. Nous avons vu la correspondance entre les équations d'espace d'état ("state space formulation") et la

fonction de transfert: celle-ci pouvant être estimée empiriquement, il apparaît possible de construire directement des modèles de dynamique des population dans le domaine fréquentiel. Cette approche, dénommée "data-based mechanistic modelling" (Young 1998), allie la puissance de la modélisation déterministe à la souplesse des outils stochastiques. La formulation d'un système d'équations correspondant à la fonction de transfert empirique permet ensuite de bénéficier directement de nombreux avantages: (i) la formulation matricielle, (ii) la dérivation directe de la solution dans le domaine temporel, (iii) l'étude aisée des effets des conditions initiales... En d'autres termes, le test d'un modèle précis s'en trouve facilité, et permettrait donc de systématiser le test d'hypothèses multiples (Hilborn and Mangel 1997).

Le terme de forçage devient ici plus explicite et perd peu à peu sa dénomination de "bruit" (*i.e.* un terme peu informatif, aléatoire et non prédictible, de peu d'intérêt écologique). Toute dynamique observée (qu'elle soit issue d'un processus géophysique ou biologique) devient potentiellement une variable explicative quelle que soit son spectre fréquentiel. Une application majeure de ces techniques porte sur le développement d'une **écologie des perturbations** dans le domaine marin. La théorie de la perturbation intermédiaire⁸ est aujourd'hui largement populaire en terrestre ou pour l'étude de la biodiversité de certains biotopes marins (de M. Huszar and Reynolds 1997). Il apparaît intéressant de généraliser cette approche au domaine fréquentiel, et de considérer non plus seulement des forçages optimaux pour des **amplitudes** intermédiaires mais aussi pour des **fréquences** intermédiaires. Le concept de "fenêtre environnementale optimale", une fois placé dans le domaine fréquentiel, n'est autre que le phénomène de **résonance** bien connu en dynamique des systèmes contrôlés. Ses conséquences, enfin, ne devraient plus être mesurées comme des indices de biodiversité, mais étudiées de multiples façons (*i.e.* mesures monospécifiques, de distribution spatiale, etc...). En d'autres termes, un comportement résonant est sans doute loin d'être une propriété anecdotique (ou une curiosité) d'un modèle, mais peut être un phénomène écologique potentiellement important. L'adéquation entre les fréquences caractéristiques de systèmes biologiques (individus, groupes d'individus, populations ou espèces) et les fréquences naturelles de leur habitat spécifique serait une piste majeure pour la compréhension (en termes évolutionnistes) de leur variabilité. Le "bruit" environnemental perd donc définitivement sa connotation négative: s'il peut effectivement camoufler certaines dynamiques ou les détruire, son rôle constructif et ordonnateur sur la biodiversité, les fluctuations ou la distribution spatiale d'espèces clés fournit un sujet de recherche passionnant et prometteur.

⁸ "the Intermediate Disturbance Theory" (Connell 1978)

Bibliographie

- Aebischer, N. J., J. C. Coulsen, and J. M. Colebrook. 1990. Parallel long-term trends across four marine trophic levels and weather. *Nature* **347**:753-755.
- Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 268-281 in N. B. Petrov and F. Csaki, editors. *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest.
- Allen, M. R., and A. W. Robertson. 1996. Distinguishing modulated oscillations from coloured noise in multivariate datasets. *Climate Dynamics* **12**:775-784.
- Alt, W. 1995. Elements of a systematic search in animal behaviour and model simulations. *Biosystems* **34**:11-26.
- Alvera-Azcarate, A., A. Barth, M. Rixen, and J. M. Beckers. 2004. Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature. *Ocean Modelling* **in press**.
- Angelini, C., D. Cava, G. Katul, and B. Vidakovic. 2004. Resampling hierarchical processes in the wavelet domain: A case study using atmospheric turbulence. RT 283/04, Georgia Institute of Technology - statistics group.
- Anonyme. 2002. North Atlantic Regional Sea Surface Temperature Product Manual. O&SI SAF Project Team Report Ocean & Sea Ice SAF.
- Arrizabalaga, H., I. Sankristobal, X. Moreno-Ventas, G. Moreno, E. Rodriguez-Martin, L. T. Kell, C. Rodriguez-Cabello, and M. Ortiz. 2002. Preliminary study on the environmental effects on bluefin catch rates in the spanish baitboat fishery of the bay of Biscay. Col. Vol. Sci. Pap. ICCAT **54**:620-640.
- Ascioti, F. A., E. Beltrami, T. O. Carroll, and C. Wirick. 1993. Is there chaos in plankton dynamics? *J. Plankton Res.* **15**:603-617.
- Atay, F. M., and Y. Altintas. 1999. Recovering smooth dynamics from time series with the aid of recurrence plots. *Physical Review E*. **59**:6593-6598.
- Auger, P., S. Charles, M. Viala, and J.-C. Poggiale. 2000. Aggregation and emergence in ecological modelling: integration of ecological levels. *Ecol. Modell.* **127**:11-20.
- Bahurel, P., P. De Mey, C. Le Provost, and P. Y. Le Traon. 2002. A GODAE prototype system with applications: Example of the Mercator system. in CNES, editor. "En route to GODAE" 2002, France.
- Bakun, A. 1996. *Patterns in the Ocean: Ocean Processes and Marine Population Dynamics*. University of California Sea Grant, San Diego, California, USA, in cooperation with Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico.
- Bard, E., G. RAISBECK, F. YIOU, and J. JOUZEL. 2000. Solar irradiance during the last 1200 years based on cosmogenic nuclides. *Tellus B* **52**:985-992.
- Barnard, J. P., C. Aldrich, and M. Gerber. 2001. Identification of dynamic process systems with surrogate data sets. *AIChE* **47**:2064-2075.
- Barnett, T. P. 1989. A solar-ocean relation: fact or fiction? *Geophysical Research Letters* **16**:803-806.
- Bartlett, M. S. 1946. On the theoretical specification of sampling properties of autocorrelated time series. *J. R. Stat. Soc.* **8**:24-411.

- Beckers, J. M., and M. Rixen. 2003. EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Oceanic Technol.* **20**:1839-1856.
- Belkin, I., and P. Cornillon. 2003. SST fronts of the Pacific coastal and marginal seas. *Pacific Oceanography* **1**:90-113.
- Belkin, I., Z. Shan, and P. Cornillon. 2004. Global pattern of ocean fronts from Pathfinder SST data: Pacific Ocean. In preparation.
- Belovsky, G. E., M. E. Ritchie, and J. Moorehead. 1989. Foraging in complex environments: when prey availability varies over time and space. *Theor. Pop. Biol.* **36**:144-160.
- Bergman, C. M., J. A. Schaefer, and L. S. N. 2000. Caribou movement as a correlated random walk. *Oecologia* **123**:364-374.
- Berryman, A., and P. Turchin. 1997. Detection of delayed density-dependence: comment. *Ecology* **78**:318-320.
- Berryman, A., and P. Turchin. 2001. Identifying the density-dependent structure underlying ecological time series. *Oikos* **92**:265-270.
- Bertino, L., G. Evensen, and H. Wackernagel. 2002. Combining geostatistics and Kalman filtering for data assimilation in an estuarine system. *Inverse Problems* **18**:1-23.
- Besag, J. 1977. Efficiency of pseudo-likelihood estimation for simple Gaussian fields. *Biometrika* **64**:616-618.
- Bider, J. R. 1968. Animal activity in uncontrolled terrestrial communities as determined by a sand transect technique. *Ecological Monographs* **38**:269-308.
- Bigelow, K. A., C. H. Boggs, and X. He. 1999. Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. *Fisheries Oceanography* **8**:178-198.
- Bjornstad, O., R. Nisbet, and J.-M. Fromentin. 2004. Trends and cohort resonant effects in age-structured populations. *J Anim Ecology* **73**:1157-1167.
- Bjørnstad, O. N., S. Champely, N. C. Stenseth, and T. Saitoh. 1996. Cyclicity and stability of grey-sided voles, *Clethrionomys rufocanus* of Hokkaido: evidence from spectral and Principal components analyses. *Philos. Trans. Roy. Soc. B* **351**:867-875.
- Bjørnstad, O. N., W. Falck, and N. C. Stenseth. 1995. A geographic gradient in small rodent density fluctuations: a statistical modelling approach. *Proc. Roy. Soc. London B* **262**:127-133.
- Bjornstad, O. N., J.-M. Fromentin, N. C. Stenseth, and J. Gjosaeter. 1999. Cycles and trends in cod populations. *Proc. Natl. Acad. Sci. USA* **96**:5066-5071.
- Bjornstad, O. N., J.-M. Fromentin, N. C. Stenseth, and J. Gjosaeter. 1999. A new test for density-dependent survival: the case of coastal cod populations. *Ecology* **80**:1278-1288.
- Block, B. A., D. P. Costa, G. W. Boehlert, and R. E. Kochevar. 2002. Revealing pelagic habitat use: the tagging of Pacific pelagics program. *Oceanologica Acta* **25**:255-266.
- Block, B. A., H. Dewar, S. B. Blackwell, T. D. Williams, E. D. Prince, C. J. Farwell, A. Boustany, S. L. H. Teo, A. Seitz, W. Andreas, and D. Fudge. 2001. Migratory Movements, Depth Preferences, and Thermal Biology of Atlantic Bluefin Tuna. *Science* **293**:1310-1314.
- Block, B. A., H. Dewar, F. C., and P. E. D. 1998. A new satellite technology for tracking the movements of Atlantic bluefin tuna. Pages 9384-9389 *in* National Academy of Sciences.

- Block, B. A., H. Dewar, T. Williams, E. Prince, C. Farwell, and F. D. 1998. Archival Tagging of Atlantic Bluefin Tuna (*Thunnus thynnus thynnus*). *Mar. Tech. Soc. J.* **32**:37-46.
- Block, B. A., and E. D. Stevens. 2001. Tuna. Physiology, ecology, and evolution. Academic Press.
- Bolviken, E., and G. Storvik. 2001. Deterministic and stochastic Particle filters in State-Space model. *in* A. Doucet and N. de Freitas, editors. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, NY.
- Boudjema, G., and B. Cazelles. 2001. Extraction of nonlinear dynamics from short and noisy time series. *Chaos, Solitons, & Fractals* **12**:2051-2069.
- Boudjema, G., and N. P. Chau. 1996. Revealing dynamics of ecological systems from natural recordings. *Ecol. Modell.* **91**:15-23.
- Boustany, A., D. J. Marcinek, J. Keen, H. Dewar, and B. A. Block. 2001. Movements and temperature preferences of Atlantic bluefin tuna (*Thunnus thynnus*) off North Carolina: A comparison of acoustic, archival and pop-up satellite tags. Pages 89-108 *in* J. R. Sibert and J. L. Nielsen, editors. *Electronic tagging and tracking in marine fisheries*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Bovet, P., and S. Benhamou. 1988. Spatial analysis of animal movements using a correlated random walk model. *J. Theor. Biol.* **131**:419-433.
- Box, G. E. P., and G. M. Jenkins. 1976. *Time series analysis, forecasting and control*. Holden Day, San Francisco.
- Boyd, I. L. 1996. Temporal scales of foraging in a marine predator. *Ecology* **77**:426-434.
- Boyd, I. L., A. Kato, and Y. Ropert-Coudert. 2004. Bio-logging science: sensing beyond the boundaries. *Mem. Natl. Inst. Polar Res.* **58**:1-14.
- Brandt, S. B. 1993. The effect of thermal fronts on fish growth: a bioenergetics evaluation of food and temperature. *Estuaries* **16**:142-159.
- Breakspear, M., M. Brammer, and P. A. Robinson. 2003. Construction of multivariate surrogate sets from nonlinear data using the wavelet transform. *Phys. D* **182**:1-22.
- Breed, M. D., R. M. Bowden, M. F. Garry, and A. L. Weicker. 1996. Giving-up time variation in response to differences in nectar volume and concentration in the giant tropical ant, *Paraponera clavata*. *J. Ins. Behav.* **9**:659-672.
- Bretherton, F., R. E. Davis, and C. Fandry. 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.* **23**:559-582.
- Brill, R., M. Lutcavage, G. Metzger, P. Bushnell, M. Arendt, J. Lucy, C. Watson, and D. G. Foley. 2002. Horizontal and vertical movements of juvenile bluefin tuna (*Thunnus thynnus*), in relation to oceanographic conditions of the western North Atlantic, determined with ultrasonic telemetry. *Fish. Bull.* **100**:155-167.
- Brillinger, D. R., and B. S. Stewart. 1996. Elephant seal movements: modelling migration. *Canadian J. Statistics* **26**:431-443.
- Brown, J. S. 1988. Patch use as an indicator of habitat preference, predation risk and competition. *Behav. Ecol. Sociobiol.* **22**:37-47.
- Brown, J. S., B. P. Kotler, and W. A. Mitchell. 1997. Competition between birds and mammals: a comparison of giving-up densities between crested larks and gerbils. *Evol. Ecol.* **11**:757-771.

- Buckland, S. T., K. B. Newman, L. Thomas, and N. B. Koesters. 2004. State-space models for the dynamics of wild animal populations. *Ecol. Modell.* **171**:157-175.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection: a practical information-theoretic approach. Springer-Verlag, New York.
- Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**.
- Carey, F. G., J. W. Kanwisher, and E. D. Stevens. 1984. Bluefin tuna warm their viscera during digestion. *Journal of Experimental Biology* **109**:1-20.
- Carey, F. G., and B. H. Robinson. 1981. *Xiphias gladius*, observed by acoustic telemetry. *Fisheries Bulletin* **79**:277-292.
- Carey, P. 1990. Les migrations: un comportement déclenché et guidé par l'environnement. *Col. Vol. Sci. Pap. ICCAT* **32**:158-168.
- Casdagli, M. C. 1989. Nonlinear prediction of chaotic time series. *Phys. D* **35**:335-336.
- Casdagli, M. C. 1997. Recurrence plots revisited. *Physica D* **108**:12-44.
- Casdagli, M. C., S. Eubank, J. D. Farmer, and J. Gibson. 1991. State-space reconstruction in the presence of noise. *Phys. D* **51**:52-98.
- Castellari, S., A. Griffa, T. M. Ozgökmen, and P.-M. Poulain. 2001. Prediction of particle trajectories in the Adriatic Sea using lagrangian data assimilation. *J. Mar. Sys.* **29**:33-50.
- Cayula, J. F., and P. Cornillon. 1995. Multi-image edge detection for SST images. *J. Atmos. Oceanic Technol.* **12**:821-829.
- Cazelles, B., and L. Stone. 2003. Detection of imperfect synchrony in an uncertain world. *J Anim Ecology* **72**:953-968.
- Chaitin, G. J. 2001. Exploring randomness. Springer-Verlag, London.
- Charnov, E. L. 1976. Optimal foraging, the marginal value theorem. *Theor. Pop. Biol.* **9**:129-136.
- Chatfield, C. 1999. The analysis of time series - an introduction. Chapman & Hall.
- Chen, J. Q., J. F. Franklin, and J. S. Lowe. 1996. Comparison of abiotic and structurally defined patch patterns in a hypothetical forest landscape. *Conservation Biology* **10**:854-862.
- Chilès, J.-P., and P. Delfiner. 1999. Geostatistics: modeling spatial uncertainty. Wiley, New York.
- Christensen, N., A. M. Bartuska, J. H. Brown, S. Carpenter, C. D'Antonio, R. Francis, J. F. Franklin, J. A. MacMahon, R. F. Noss, D. J. Parsons, C. H. Peterson, M. G. Turner, and R. G. Woodmansee. 1996. The report of the Ecological Society of America committee on the scientific basis for ecosystem management. *Ecological Applications* **6**:665-691.
- Clark, D. R., K. D. Aazem, and G. C. Hays. 2001. Zooplankton abundance and community structure over a 4000 km transect in the north-east Atlantic. *J. Plankton Res.* **23**:365-372.
- Cliff, A. D., and J. K. Ord. 1981. Spatial Processes. Pion Press, London.
- Cohen, J. E. 1994. Marine and continental foodwebs: three paradoxes? *Philosophical Transactions of the Royal Society of London B* **343**:57-68.

- Collette, B. B. 1979. Adaptations and systematics of the mackerels and tunas. *in* G. D. Sharp and A. E. Dizon, editors. The physiological ecology of tunas. Academic Press, New York.
- Connell, J. H. 1978. Diversity in tropical rain forest and coral reefs. *Science* **199**:1302-1310.
- Cort, J. L. 1991. Age and growth of the bluefin tuna *Thunnus thynnus* (L.) of the Northeast Atlantic. Col. Vol. Sci. Pap. ICCAT **35**:214-230.
- Coutant, C. C. 1977. Compilation of temperature preference data. *Journal of the Fisheries Research Board of Canada* **34**:739-745.
- Crane, J. 1936. Notes on the biology and ecology of giant tuna *Thunnus thynnus*, L., observed at Portland, Maine. *Zoologica* **212**:207-212.
- Crane, J. 1936. *Thunnus thynnus*, L., observed at Portland, Maine. *Zoologica* **212**:207-212.
- Cressie, N. 1993. Statistics for spatial data. Wiley, New York.
- Cronin, T. W., and N. Shashar. 2001. The linearly polarized light field in clear, tropical marine waters: spatial and temporal variation of light intensity, degree of polarization and e-vector angle. *J. Exp. Biol.* **204**.
- Crowder, L. B. 1985. Optimal foraging and feeding mode shifts in fishes. *Environ. Biol. Fish* **12**:57-62.
- Crowley, T. J. 2000. Causes of climate change over the past 1000 years. *Science* **289**:270-277.
- Cubasch, U., R. Voss, G. C. Hegerl, J. Waszkewitz, and T. J. Crowley. 1997. Simulation of the influence of solar radiation variations on the global climate with an ocean-atmosphere circulation model. *Climate Dynamics* **13**:757-767.
- Cury, P., O. Anneville, F. X. Bard, A. Fonteneau, and C. Roy. 1996. Obstinate north atlantic bluefin tuna (*Thunnus thynnus thynnus*): an evolutionary perspective to consider spawning migration. Pages 239-249 *in* ICCAT 25th Symposium.
- Cury, P., and C. Roy. 1989. Optimal environmental window and pelagic fish recruitment success in upwelling areas. *Can. J. Fish. Aquat. Sci.* **46**:670-680.
- Cushing, D. H. 1982. Climate and fisheries. Academic Press, London, UK.
- Cushing, D. H. 1990. Plankton production and year class strength in fish populations: an update of the match-mismatch hypothesis. *Adv. Mar. Biol.* **26**:249-293.
- Cushing, D. H., and R. R. Dickson. 1976. The biological response in the sea to climatic changes. *Adv. Mar. Biol.* **14**:1-122.
- Dagorn, L., F. Menczer, P. Bach., and R. J. Olson. 2000. Co-evolution of movement behaviours by tropical pelagic predatory fishes in response to prey environment: a simulation model. *Ecol. Modell.* **134**:325-341.
- Dahia, K. 2005. Nouvelles méthodes en filtrage particulière: application au recalage de navigation inertielle par mesures altimétriques. PhD. Univ. Joseph Fourier.
- Dale, M. R. T., P. Dixon, M.-J. Fortin, P. Legendre, D. E. Myers, and M. S. Rosenberg. 2002. Conceptual and mathematical relationships among methods for spatial analysis. *Ecography* **25**:558-577.
- de M. Huszar, V. L., and C. S. Reynolds. 1997. Phytoplankton periodicity and sequences of dominance in an Amazonian flood-plain lake (Lago Batata, Pará, Brasil): responses to gradual environmental change. *Hydrobiologia* **346**:169-181.

- De Metrio, G., G. P. Arnold, B. A. Block, J. M. De la Serna, M. Deflorio, M. Cataldo, C. Yannopoulos, P. Megalofonou, S. Beemer, C. Farwell, and A. Seitz. 2002. Behaviour of post-spawning atlantic bluefin tuna tagged with pop-up satellite tags in the mediterranean and eastern atlantic. *Col. Vol. Sci. Pap. ICCAT* **54**:415-424.
- De Metrio, G., G. P. Arnold, J. M. De la Serna, C. Yannopoulos, P. Megalofonou, A. A. Buckley, and M. Pappalepore. 2001. Further results of tagging mediterranean bluefin tuna with pop-up satellite-detected tags. *Col. Vol. Sci. Pap. ICCAT* **52**:776-783.
- DeAngelis, D. L., and G. T. Yeh. 1984. An introduction to modeling migratory behaviour of fishes. Pages 445-469 *in* J. D. McCleave, G. P. Arnold, J. J. Dodson, and W. H. Neill, editors. *Mechanisms of migration in fishes*. Plenum Publishing Corporation.
- Dempster, A., N. M. Laird, and D. B. Rubin. 1977. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B* **39**:1-38.
- Deutschlander, M. E., J. B. Phillips, and S. C. Borland. 1999. The case for light-dependent magnetic orientation in animals. *J. Exp. Biol.* **202**:891-908.
- Dieckmann, U., and R. Law. 1999. Relaxation projection and the method of moments. Interim Report 040, International Institute for Applied Systems Analysis.
- Diehl, S. F., and J. Wells Budd. 2000. Geographic window sizes applied to remote-sensing sea-surface temperature front detection. Michigan Technological University, Houghton.
- Dippner, J. W., R. Heerkloss, and J. P. Zbilut. 2002. Recurrence Quantification Analysis as a tool for characterisation of non-linear mesocosm dynamics. *Mar. Ecol. Prog. Ser.* **242**:29-37.
- Doney, S., D. M. Glover, S. J. McCue, and M. Fuentes. 2004. Mesoscale variability of SeaWiFS satellite ocean color: global patterns and spatial scales. *Journal of Geophysical Research* **Submitted**.
- Donoho, D. 1992. Nonlinear Wavelet Methods for Recovering Signals, Images, and Densities from indirect and noisy data. Technical Report 403, Stanford University, Stanford, CA.
- Doucet, A., and S. Godsill. 1998. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. Technical report CUED/F-INFENG/TR. 310, Cambridge University.
- Doumenge, F. 1998. L'histoire des pêcheries thonières. *Col. Vol. Sci. Pap. ICCAT* **50**:753-802.
- Dufour, P., and J.-M. Stretta. 1973. Fronts thermiques et thermohalins dans la région du Cap Lopez (Golfe de Guinée) en juin-juillet 1972: phytoplancton, zooplancton, micronecton et pêche thonière. *Doc. Sci. Cent. Rech. Océanogr. Abidjan ORSTOM* **4**:99-142.
- Eckmann, J.-P., S. O. Kamphorst, and D. Ruelle. 1987. Recurrence Plots of Dynamical Systems. *Europhysics Letters* **5**:973-977.
- Eckmann, J.-P., and D. Ruelle. 1985. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys* **57**:617.
- Eckmann, J.-P., and D. Ruelle. 1992. Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Phys. D* **56**:185-187.
- Engen, S., and R. Lande. 1996. Population dynamic models generating the lognormal species abundance distribution. *Mathematical Biosciences* **132**:169-183.
- Etienne, H., and I. Charon. 2004. Dossier de validation PSY2v1 trimestriel n°4. MOO-RP-410-264-MER, Groupe Mission Mercator, Ramonville St-Agne.

- Etnoyer, P., D. Canny, B. Mate, and L. Morgan. 2004. Persistent Pelagic Habitats in the Baja California to Bering Sea Ecoregion. *Oceanogr.* **17**:90-101.
- Evans, G. T., and J. S. Parslow. 1985. A model of annual plankton cycles. *Biol. Oceanogr.* **3**:327-347.
- Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods for forecast error statistics. *J. Geophys. Res.* **99**:10143-10162.
- Evensen, G. 2003. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics* **53**:343-367.
- Evensen, G., and P. J. van Leeuwen. 2000. An Ensemble Kalman smoother for Nonlinear Dynamics. *Mon. Weath. Rev.* **128**:1852-1867.
- Fagan, W. F., M.-J. Fortin, and C. Soykan. 2003. Integrating edge detection and dynamic modeling in quantitative analyses of ecological boundaries. *Bioscience* **53**:730-738.
- Farrugio, H. 1981. Exploitation et dynamique des populations de thon rouge, *Thunnus thynnus* (Linné 1758) Atlanto-Méditerranéennes. PhD dissertation. Université des Sciences et Techniques du Languedoc.
- Fearnhead, P. 1998. Sequential Monte Carlo methods in filter theory. PhD thesis. University of Oxford.
- Fedorov, K. N. 1986. The physical nature and structure of oceanic fronts. Springer-Verlag, Berlin.
- Feller, W. 1968. An introduction to probability theory and its applications. Wiley, New York.
- Flierl, G., D. Grünbaum, S. Levin, and D. B. Olson. 1999. From Individuals to Aggregations: the Interplay between Behavior and Physics. *J. Theor. Biol.* **196**:397-454.
- Foldager, L., and J. Pedersen. 2002. Spatial Distribution of Detected Single Fish: An Application of the Log Gaussian Cox Process. 03, Biometry Research Unit, Department of Agricultural Systems, Danish Institute of Agricultural Sciences.
- Fonteneau, A. 1997. Atlas des pêcheries thonières tropicales. Captures mondiales et environnement. ORSTOM editions, Paris.
- Fortin, M.-J., M. R. T. Dale, and J. Ver Hoef. 2002. Spatial analysis in ecology. Pages 2051-2058 in A. H. El-Shaarawi and W. W. Piegorsch, editors. *Encyclopedia of environmetrics*. Wiley.
- Fortin, M.-J., and G. M. Jacquez. 2000. Randomization tests and spatially autocorrelated data. *Bulletin of the Ecological Society of America* **81**:201-205.
- Fournier, R. O., M. Van Det, N. B. Hargreaves, J. S. Wilson, T. A. Clair, and R. Ernst. 1984. Physical factors affecting summer distribution of chlorophyll a off southwestern Nova Scotia. *Limnology and Oceanography* **29**:517-526.
- Franks, P. J. S. 1992. Swim or sink: accumulation of biomass at fronts. *Mar. Ecol. Prog. Ser.* **82**:1-12.
- Fretwell, S., and H. Lucas. 1970. On the territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheoretica* **19**:16-36.
- Fromentin, J.-M. 2001. Can stochastic variations in recruitment induce long-term fluctuations in the carrying capacity? *Communication ICCAT SCRS/01/34*.

- Fromentin, J.-M. 2002. Final Report of STROMBOLI. EU-DG XIV project 99/022, European Community - DG XIV, Brussels.
- Fromentin, J.-M. 2003. The East Atlantic and Mediterranean Bluefin tuna management: uncertainties and alternatives. *Scientia Marina* **67**:51-62.
- Fromentin, J.-M., and A. Fonteneau. 2001. Fishing effects and life history traits: a case study comparing tropical versus temperate tunas. *Fisheries Research* **53**:133-150.
- Fromentin, J.-M., A. Fonteneau, and H. Farrugio. 2000. Biological key reference points and natural long-term fluctuations: the case of the Atlantic bluefin tuna. Col. Vol. Sci. Pap. ICCAT **51**:2072-2084.
- Fromentin, J.-M., R. A. Myers, O. N. Bjornstad, N. C. Stenseth, J. Gjosaeter, and C. Hartvig. 2001. Effects of density-dependent and stochastic processes on the regulation of cod populations. *Ecology* **82**:567-579.
- Fromentin, J.-M., and B. Planque. 1996. *Calanus* and environment in the eastern North Atlantic. 2) Role of the North Atlantic Oscillation on *Calanus finmarchicus* and *C. helgolandicus*. *Mar. Ecol. Prog. Ser.* **134**:111-118.
- Gao, J., and H. Cai. 2000. On the structures and quantification of recurrence plots. *Physics Letters A* **270**:75-87.
- Garcia, A., F. Alemany, P. Velez-Belchí, J. L. López Jurado, J. M. de la Serna, C. González Pola, J. M. Rodriguez, and J. Jansá. 2003. Bluefin tuna and associated species spawning grounds in the oceanographic scenario of the Balearic archipelago during June 2001. *Collective Volume of Scientific Papers ICCAT* **56**:138-148.
- Gibson, J., and D. E. Myers. 2003. A meta-analysis of the habitat carrying capacity and maximum reproductive rate of anadromous alewife in eastern north America. *American Fisheries Society Symposium* **35**:211-221.
- Gisiger, T. 2001. Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biol. Rev.* **76**:161-209.
- Giuliani, A., R. Benigni, and J. P. Zbilut. 2000. Nonlinear methods in the analysis of protein sequences: a case study in rubredoxins. *Biophys. J.* **78**:136-148.
- Gordon, N., D. J. Salmond, and A. F. M. Smith. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings* **2**:107-113.
- Gosz, J. R. 1993. Ecotone hierarchies. *Ecological Applications* **3**:369-376.
- Gower, J. F. R., K. Denman, and R. J. Holyer. 1980. Phytoplankton patchiness indicates the fluctuation spectrum of mesoscale oceanic structure. *Nature* **288**:157-159.
- Grassberger, P., T. Schreiber, and C. Schaffrath. 1991. Non-linear time sequence analysis. *Int. J. Bifurcation and Chaos* **1**:521.
- Greatbach, R. J., A. F. Fanning, and A. D. Goulding. 1991. A diagnosis of interpentadal circulation changes in the North Atlantic. *J. Geophys. Res.* **96**:22009-20023.
- Greiner, E., Y. Drillet, R. Bourdalle-Badie, J.-M. Lellouche, and N. Verbrugge. 2004. PSY2v1: rejeu B003, analyse de septembre 2002 à août 2003. MOO-NT-410-273-MER, Groupe Mission Mercator, Ramonville Saint-Agne.
- Gronnevik, R., and G. Evensen. 2001. Application of ensemble-based techniques in fish stock assessment. *Sarsia* **86**:517-526.

- Gunn, J. S., T. Polacheck, T. L. Davis, M. Sherlock, and A. Betlehem. 1994. The application of archival tags to study the movement, behaviour and physiology of southern bluefin tuna, with comments on the transfer of the technology to groundfish research. 21, ICES.
- Gurney, W. S., and R. Nisbet. 1998. *Ecological dynamics*. Oxford University Press, Oxford.
- Gurney, W. S., D. C. Speirs, S. N. Wood, E. D. Clarke, and M. Heath. 2001. Simulating spatially and physiologically structured populations. *Journal of Animal Ecology* **70**:881-894.
- Gutenkunst, R., N. K. Newlands, M. Lutcavage, and L. Edelstein-Keshet. 2004. Inferring resource distributions from Atlantic bluefin tuna movements: an analysis based on net displacement and length of track. *Ecology*.
- Haight, J. D. 1996. The impact of solar variability on climate. *Science* **272**:981-984.
- Halley, J. 1996. Ecology, evolution and 1/f noise. *Tree* **11**:33-37.
- Harvey, A. C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.
- Hastings, A., C. L. Horn, S. P. Ellner, P. Turchin, and H. C. J. Godfray. 1993. Chaos in ecology: is mother nature a strange attractor? *Annu. Rev. Ecol. Syst.* **25**:1-33.
- Hastings, A., and T. M. Powell. 1991. Chaos in a three-species food chain. *Ecology* **73**:896-903.
- Haury, L. R., J. A. Mc Gowan, and P. H. Wiebe. 1978. Patterns and processes in the time-space scales of plankton distributions. Pages 277-327 *in* J. H. Steele, editor. *Spatial pattern in plankton communities*. Plenum Press, New-York.
- Hegger, R., L. Kantz, L. Matassimi, and T. Schreiber. 2000. Coping with non-stationarity by overembedding. *Phys. Rev. Lett.* **84**:4092-4095.
- Hensen, V. 1887. *Über die Bestimmung des Planktons oder des im Meere treibenden Materials an Pflanzen und Thieren. in Kommission zur wissenschaftlichen Untersuchungen der deutschen Meere in Kiel, 1882-1886.*
- Heyen, H., H. Fock, and W. Greve. 1998. Detecting relationships between the interannual variability in ecological time series and climate using a multivariate statistical approach-- a case study on Helgoland Roads zooplankton. *Climate Research* **10**:179-191.
- Higgins, C. L., and R. E. Strauss. 2004. Discrimination and classification of search paths produced by different search-tactics models. *Behav. Ecol.* **in press**.
- Hilborn, R., and M. Mangel. 1997. *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton.
- Hjort, J. 1914. Fluctuations in the great fisheries of northern Europe. *Rapp. P.-V. Reun. Cons. Int. Explor. Mer* **20**:1-227.
- Holbrook, N., and N. Bindoff. 2000. A statistically efficient mapping technique for four-dimensional ocean temperature data. *J. Atmos. Oceanic Technol.* **17**:831-846.
- Holyst, J. A., M. Zebrowska, and K. Urbanowicz. 2001. Observations of deterministic chaos in financial time series by recurrence plots, can one control chaotic economy? *The European Physical Journal B* **20**:531-535.
- Horvath, G., and D. Varju. 1995. Underwater Refraction--Polarization Patterns of Skylight Perceived by Aquatic Animals Through Snell's Window of the Flat Water Surface. *Vision Research* **35**:1651-1666.

- Hoyt, D. V., and K. H. Schatten. 1993. A discussion of plausible solar irradiance variations 1700-1992. *J. Geophys. Res.* **98**:18895-18906.
- Hubel, H., C. Wolfe, and L.-A. Meyer-Reil. 1998. Salinity, inorganic nutrients and primary production in a shallow coastal inlet in the southern Baltic Sea: results from long-term observations (1960-1989). *International Review of Hydrobiology* **83**:479-499.
- Huisman, J., and F. J. Weissing. 1999. Biodiversity of plankton by species oscillations and chaos. *Nature* **402**:407-410.
- Hurrell, J. W. 1995. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitations. *Science* **269**:676-679.
- Hurrell, J. W., Y. Kushnir, and M. Visbeck. 2001. The North Atlantic Oscillation. *Science* **291**:603-605.
- Hutchinson, G. E. 1961. The paradox of the plankton. *Am. Nat.* **95**:137-145.
- ICCAT. 2002. Report of the ICCAT Workshop on environment and tuna recruitment (Madrid, Spain, May 7-12, 2001). *Collective Volume of Scientific Papers ICCAT* **54**:895-952.
- Inagake, D., H. Yamada, K. Segawa, M. Okazaki, A. Nitta, and T. Itoh. 2001. Migration of young Bluefin tuna, *Thunnus orientalis*, through archival tagging experiments and its relation with oceanographic conditions in the Western North Pacific. *Bulletin of the National Research Institute of Far Seas Fisheries* **38**:53-81.
- Iverson, R. L. 1990. Control of marine fish production. *Limnology and Oceanography* **35**:1593-1604.
- Iwanski, J. S., and E. Bradley. 1998. Recurrence plot analysis: To embed or not to embed? *Chaos* **8**:861.
- Jiang, H., J. Eiriksson, M. Schulz, K.-L. Knudsen, and M.-S. Seidenkrantz. 2005. Evidence for solar forcing of sea-surface temperature on the North Icelandic Shelf during the late Holocene. *Geology* **33**:73-76.
- Jiang, L., and N. Shao. 2004. Red environmental noise and the appearance of delayed density-dependence in age-structured populations. *Proc. R. Soc. Lond. B* **271**:1059-1064.
- Joergensen, S. E. 1997. *Integration of ecosystem theory: a pattern*. Kluwer Academic Publishers, Dordrecht.
- Johns, B., P. Marsaleix, C. Estournel, and R. Véhil. 1992. On the wind-driven coastal upwelling in the Gulf of Lions. *J. Mar. Syst.* **3**:309-320.
- Jones, P. D. 1994. Hemispheric surface air temperature variations: a reanalysis and update to 1993. *J. Climate* **7**:1793-1802.
- Jonzén, N., P. Lundberg, E. Ranta, and V. Kaitala. 2002. The irreducible uncertainty of the demography-environment interaction in ecology. *Proc. R. Soc. Lond. B* **269**:221-225.
- Kacelnik, A., J. R. Krebs, and C. Bernstein. 1992. The ideal free distribution and predator-prey populations. *Trends in Ecology & Evolution* **7**:50-55.
- Kaitala, V., P. Lundberg, J. Ripa, and J. Ylikarjula. 1997. Red, Blue and Green: Dyeing Population Dynamics. *Ann. Zool. Fennici* **34**:217-228.
- Kaitala, V., J. Ylikarjula, E. Ranta, and P. Lundberg. 1997. Population Dynamics and the Colour of Environmental Noise. *Proc. R. Soc. Lond. B* **264**:943-948.

- Kalman, R. 1960. A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME--Journal of Basic Engineering **82**:35-45.
- Kamenkovich, V. M., M. N. Koshlyakov, and A. S. Monin. 1987. Synoptic eddies in the ocean. Godrometeoizdat, Leningrad.
- Kantz, H., and T. Schreiber. 1997. Nonlinear Time Series Analysis. Cambridge University Press, Cambridge.
- Karakulak, S., I. Oray, A. Corriero, M. Deflorio, N. Santamaria, S. Desantis, and G. De Metrio. 2004. Evidence of a spawning area for the bluefin tuna (*Thunnus thynnus* L.) in the eastern Mediterranean. J Appl Ichthyol **20**:318-320.
- Kareiva, P. M., and N. Shigesada. 1983. Analyzing insect movement as a correlated random walk. Oecologia **56**:234-238.
- Kazmin, A. S., and M. M. Rienecker. 1996. Variability and frontogenesis in the large-scale oceanic frontal zones. J. Geophys. Res. **101**:907-921.
- Kendall, B. E., C. J. Briggs, W. W. Murdoch, P. Turchin, S. P. Ellner, E. McCauley, R. M. Nisbet, and S. N. Wood. 1999. Why do populations cycle? A synthesis of statistical and mechanistic modelling approaches. Ecology **80**:1789-1805.
- Kendall, D. G. 1949. Stochastic processes and population growth. J. R. Stat. Soc. **11**:230-264.
- Kimura, S., M. Nakai, and T. Sugimoto. 1997. Migration of albacore, *Thunnus alalunga*, in the North Pacific Ocean in relation to large oceanic phenomena. Fish. Oceanogr. **6**:51-57.
- Kinder, T. H., G. L. Hunt, D. C. Schneider, and J. D. Schumacher. 1983. Oceanic fronts around the Pribilof islands, Alaska: correlation with seabirds. Estuarine, coastal and shelf science **16**:309-319.
- Kirby, D. S., E. R. Abraham, M. J. Uddstrom, and H. Dean. Submitted. Spatial scales of tuna aggregations in surface longline data. New Zealand Journal of Marine and Freshwater Research.
- Kirby, D. S., Ø. Fiksen, and P. J. B. Hart. 2000. A dynamic optimization model for the behaviour of tunas at ocean fronts. Fish. Oceanogr. **9**:328-342.
- Klees, R., P. Ditmar, and P. Broersen. 2002. How to handle colored observation noise in large least-squares problems. Journal of Geodesy **76**:629-640.
- Knoppien, P., and J. Reddinguis. 1985. Predators with two modes of searching: a mathematical model. J. Theor. Biol. **114**:273-301.
- Kohn, R., and C. F. Ansley. 1986. Estimation, prediction, and interpolation for ARIMA models with missing data. J. Am. Stat. Assoc. **81**:751-761.
- Korsmeyer, K. E., and H. Dewar. 2001. Tunas metabolism and energetics. *in* B. A. Block and E. D. Stevens, editors. Tuna: physiology, ecology and evolution. Academic Press, New York.
- Krakauer, D. C., and M. A. Rodriguez-Girones. 1995. Searching and learning in a random environment. J. Theor. Biol. **177**:417-419.
- Laakso, J., V. Kaitala, and E. Ranta. 2001. How does environmental variation translate into biological processes? Oikos **92**:119-122.
- Largier, J. L. 1993. Estuarine fronts: how important are they? Estuaries **16**:1-11.

- Laurs, R. M., P. C. Fiedler, and D. R. Montgomery. 1984. Albacore tuna catch distribution relative to environment features observed from satellites. *Deep Sea Research* **31**:1085-1099.
- Laurs, R. M., and J. R. Lynn. 1977. Seasonal migration of North Pacific albacore, *Thunnus alalunga*, into North American coastal waters: distribution, relative abundance and association with transition zone waters. *Fish. Bull.* **75**:795-822.
- Law, R., and U. Dieckmann. 1999. Moment approximations of individual-based models. Interim Report 043, International Institute for Applied Systems Analysis.
- Le Fevre, J. 1986. Aspects of the biology of frontal systems. *Adv. Mar. Biol.* **23**:163-299.
- Le Traon, P. Y., F. Nadal, and N. Ducet. 1997. An improved mapping method of multisatellite altimeter data. *J. Atmos. Oceanic Technol.* **15**:522-534.
- Lean, J., and D. Rind. 1998. Climate forcing by changing solar radiation. *J. Climate* **11**:3069-3094.
- Legendre, L., and S. Demers. 1984. Towards dynamic biological oceanography and limnology. *Can. J. Fish. Aquat. Sci.* **41**:2-19.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**:1659-1673.
- Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetation* **80**:107-138.
- Legendre, P., and L. Legendre. 1998. *Numerical Ecology*. Elsevier, Amsterdam, The Netherlands.
- Legendre, P., R. R. Sokal, N. L. Oden, A. Vaudor, and J. H. Kim. 1990. Analysis of variance with spatial autocorrelation in both the variable and the classification criterion. *Journal of Classification* **7**:53-75.
- Lehodey, P., J.-M. Andre, M. Bertignac, J. Hampton, A. Stoens, C. Menkes, L. Memery, and N. Grima. 1998. Predicting skipjack tuna forage distributions in the equatorial Pacific using a coupled dynamical bio-geochemical model. *Fisheries Oceanography* **7**:317-325.
- Levin, S. A., and L. A. Segel. 1976. Hypothesis for origin of planktonic patchiness. *Nature* **259**:659.
- Lévy, M., P. Klein, and A.-M. Tréguier. 2001. Impact of sub-mesoscale physics on production and subduction of phytoplankton in an oligotrophic regime. *Journal of Marine Research* **59**:535-565.
- Lewis, F. L. 1986. *Optimal estimation with an introduction to stochastic control theory*. Wiley.
- Lindahl, O., and L. Hernroth. 1988. Large-scale and long-term variations in the zooplankton community of the Gullmar fjord, Sweden, in relation to advective processes. *Mar. Ecol. Prog. Ser.* **43**:161-171.
- Liorzou, B. 1989. Espèces et pêcheries concernées dans les eaux tempérées. Pages 3-8 in J.-Y. Le Gall, editor. *Téledétection satellitaire et pêcheries thonières océaniques*. FAO, Rome.
- Lohmann, G., N. Rambu, and M. Dima. 2004. Climate signature of solar irradiance variations: analysis of long-term instrumental, historical and proxy data. *International Journal of Climatology* **24**:1045-1056.
- Longhurst, A. R. 1995. Seasonal cycles of pelagic production and consumption. *Progress In Oceanography* **36**:77-167.
- Longhurst, A. R. 1998. *Ecological geography of the sea*. Academic press, San Diego.

- Lovejoy, S., W. J. S. Currie, Y. Tessier, M. R. Claereboudt, B. E., J. C. Roff, and D. Schertzer. 2001. Universal multifractals and ocean patchiness: phytoplankton, physical fields and coastal heterogeneity. *JOURNAL OF PLANKTON RESEARCH* **23**:117-141.
- Lundberg, P., E. Ranta, J. Ripa, and V. Kaïtala. 2000. Population variability in space and time. *Trends in Ecology and Evolution* **15**:460-464.
- Lutcavage, M., R. Brill, G. B. Skomal, C. C. Bradford, and P. W. Howey. 1999. Results of pop-up satellite tagging of spawning size class fish in the Gulf of Maine: do North Atlantic bluefin tuna spawn in the mid-Atlantic? *Can. J. Fish. Aquat. Sci.* **56**:173-177.
- Lutcavage, M., R. Brill, G. B. Skomal, B. C. Chase, J. Goldstein, and J. Tutein. 2000. Tracking adult North Atlantic bluefin tuna (*Thunnus Thynnus*) in the northwestern Atlantic using ultrasonic telemetry. *Mar. Biol.* **137**:347-358.
- Lutcavage, M., J. Goldstein, and S. Kraus. 1996. Distribution, relative abundance, and behavior of giant bluefin tuna in New England waters. *Col. Vol. Sci. Pap. ICCAT* **46**:332-347.
- Mac Call, A. D. 1990. *Dynamic geography of marine fish populations*. Univ. of Washington Press.
- Mackas, D. L., L. Washburn, and S. L. Smith. 1991. Zooplankton community pattern associated with a California Current cold filament. *J. Geophys. Res.* **96**:14781-14792.
- Magnuson, J. J. 1973. Comparative studies of adaptations for continuous swimming and hydrostatic equilibrium of scombroid and xiphoid fishes. *Fish. Bull.* **71**:337-356.
- Magnuson, J. J. 1978. Locomotion by scombrid fishes: hydromechanics, morphology and behavior. Pages 239-313 *in* W. S. Hoar and D. J. Randall, editors. *Fish physiology*. Academic Press, New York.
- Mahadevan, A., and J. W. Campbell. 2002. Biogeochemical patchiness at the sea surface. *Geophysical Research Letters* **29**:32-36.
- Mahadevan, A., and J. W. Campbell. 2003. Biogeochemical Variability at the Sea Surface: How it is Linked to Process Response Times. Pages 215-227 *in* *A Handbook of Scales in Aquatic Ecology*. CRC Press.
- Mandelbrot, B. B., and J. R. Wallis. 1969. Some long-run properties of geophysical records. *Water Resource Res.* **5**:321-340.
- Marsac, F. 1999. Changements climatiques observés dans l'Atlantique depuis les années 50 et impacts possibles sur quelques stocks de thon et leur exploitation. *Col. Vol. Sci. Pap. ICCAT* **49**:346-370.
- Martin, A. P., K. J. Richards, A. Bracco, and A. Provenzale. 2002. Patchy productivity in the open ocean. *Global Biogeochemical Cycles* **16**:1-9.
- Marwan, N., M. H. Trauth, M. Vuille, and J. Kurths. 2003. Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods. *Climate Dynamics* **21**:317-326.
- Marwan, N., N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths. 2002. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E.* **66**:026702.
- Marwan, N., N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths. 2002. Recurrence Plot Based Measures of Complexity and its Application to Heart Rate Variability Data. *Physical Review E.* **66**:026702.

- Mather, F. J., J. M. Mason, and A. C. Jones. 1995. Historical document: life history and fisheries of Atlantic bluefin tuna. NOAA Technical Memorandum 370, NMFS-SEFSC.
- Matheron, G., editor. 1971. The Theory of Regionalized Variables and Its Applications, Fontainebleau.
- Maul, G. A., F. Williams, M. Roffer, and F. M. Sousa. 1984. Remotely sensed oceanographic patterns and variability of bluefin tuna (*Thunnus thynnus*) catches in the Gulf of Mexico. *Oceanologica Acta* **7**:469-480.
- Maury, O., D. Gascuel, F. Marsac, A. Fonteneau, and A.-L. De Rosa. 2001. Hierarchical interpretation of non-linear relationships linking yellowfin tuna (*Thunnus albacares*) distribution to the environment in the Atlantic Ocean. *Can. J. Fish. Aquat. Sci.* **58**:1-13.
- May, R. M., and G. F. Oster. 1976. Bifurcation and dynamic complexity in simple ecological models. *Am. Nat.* **110**:573-599.
- McCann, K., L. W. Botsford, and A. Hasting. 2003. Differential response of marine populations to climate forcing. *Can. J. Fish. Aquat. Sci.* **60**:971-985.
- McCann, K., and P. Yodzis. 1994. Biological conditions for chaos in a three-species food chain. *Ecology* **75**:561-564.
- McGowan, J. A., D. R. Cayan, and L. M. Dorman. 1998. Climate-ocean variability and ecosystem response in the Northeast Pacific. *Science* **281**:210-217.
- Medvinsky, A. B., S. V. Petrovskii, I. A. Tikhonova, E. Venturino, and H. Malchow. 2001. Chaos and regular dynamics in model multi-habitat plankton-fish communities. *J. Biosci.* **1**.
- Meyer, R., and N. Christensen. 2001. Fast bayesian reconstruction of chaotic dynamical systems via extended Kalman filtering. *Physical Review E.* **65**:1-8.
- Miller, P. 2000. Multispectral front maps for auto-detection of ocean colour features from SeaWiFS and MODIS. *in* Oceans from Space, Venice.
- Millot, C. 1990. The Gulf of Lion's hydrodynamics. *Continental Shelf Res.* **10**:143-149.
- Millot, C. 1999. Circulation in the western mediterranean sea. *Journal of Marine Systems* **20**:423-442.
- Milotti, E. 1995. Linear processes that produce 1/f or flicker noise. *Phys. Rev. E.* **4**:3087-3103.
- Milotti, E. 2001. 1/f noise: a pedagogical review.
- Morales, J. M., D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology* **85**:2436-2445.
- Morin, E. 1990. Introduction à la pensée complexe. Editions ESF.
- Munk, P., P. O. Larsson, D. Danielsen, and E. Moksness. 1995. Larval and small juvenile cod *Gadus morhua* concentrated in the highly productive areas of a shelf break front. *Mar. Ecol. Prog. Ser.* **125**:21-30.
- Musyl, M. K., R. Brill, D. S. Curran, J. S. Gunn, J. R. Hartog, R. D. Hill, D. W. Welch, J. P. Eveson, C. H. Boggs, and R. E. Brainard. 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. *Reviews in Fish Biology and Fisheries* **In press**.

- Myers, R. A., B. R. MacKenzie, K. G. Bowen, and N. J. Barrowman. 2001. What is the carrying capacity of fish in the ocean? A meta-analysis of population dynamics of North Atlantic cod. *Can. J. Fish. Aquat. Sci.* **58**:2404-2419.
- Myers, R. A., and G. Mertz. 1998. Reducing uncertainty in the biological basis of fisheries management by meta-analysis of data from many populations; a synthesis. *Fish. Res.* **37**:51-60.
- Nakamura, H., and A. S. Kazmin. 2003. Decadal changes in the North Pacific oceanic frontal zones as revealed in ship and satellite observations. *J. Geophys. Res.* **103**:85.
- Newlands, N. K., M. Lutcavage, and T. J. Pitcher. 2004. Analysis of foraging movements of bluefin tuna (*Thunnus thynnus*). Individuals switch between two modes of search behavior. *Population Ecology* **46**:39-53.
- Nielsen, A. 2004. Estimating fish movement. Royal Veterinary and Agricultural University, Denmark.
- Nisbet, R., and W. S. Gurney. 1982. Modelling fluctuating populations. Wiley, Chichester.
- Nixon, S. W. 1988. Physical energy inputs and the comparative ecology of lake and marine ecosystems. *Limnology and Oceanography* **33**:1005-1025.
- Olson, D. B., G. L. Hitchcock, A. J. Mariano, C. J. Ashjian, G. Peng, R. W. Nero, and G. P. Podesta. 1994. Life on the edge: marine life and fronts. *Oceanogr.* **7**:52-60.
- Olson, D. B., and G. P. Podesta. 1986. Oceanic fronts as pathways in the sea. Pages 1-14 in W. F. Herrnkind and A. B. Thistle, editors. Signposts in the sea, a multidisciplinary workshop on marine animal orientation and migration, Florida State University, Tallahassee, Florida.
- Ortega, G., and L. Romanelli. 1995. Detecting hidden frequencies in dynamical time series: A numerical report. *Chaos, Solitons & Fractals* **6**:411-415.
- Ottersen, G., B. Planque, A. Belgrano, E. Post, P. C. Reid, and N. C. Stenseth. 2001. Ecological effects of the North Atlantic Oscillation. *Oecologia* **128**:1-14.
- Ozgökmen, T. M., A. Griffa, and A. J. Mariano. 2000. On the predictability of lagrangian trajectories in the ocean. *J. Atmos. Oceanic Technol.* **17**:366-383.
- Özgökmen, T. M., L. I. Piterbarg, A. J. Mariano, and E. H. Ryan. 2001. Predictability of drifter trajectories in the tropical Pacific Ocean. *J. Phys. Oceanogr.* **31**:2691-2720.
- Papavasiliou, A. 2002. Adaptive particle filters with applications. PhD thesis. University of Princeton.
- Parker, D. E., T. P. Legg, and C. K. Folland. 1992. A new daily Central England Temperature series, 1772--1991. *International Journal of Climatology* **12**:317-342.
- Parkyn, D. C., J. D. Austin, and C. W. Hawryshyn. 2003. Acquisition of polarized-light orientation in salmonids under laboratory conditions. *Animal Behaviour* **65**:893-904.
- Parsons, T. R. 1979. Some ecological, experimental and evolutionary aspects of the upwelling ecosystem. *South African Journal of Science* **75**:536-540.
- Pascual, M., and S. P. Ellner. 2000. Linking ecological patterns to environmental forcing via nonlinear time series modeling. *Ecology* **81**:2767-2780.

- Patil, D. A. S., B. R. Hunt, and J. A. Carton. 2000. Identifying Low-Dimensional Nonlinear Behavior in Atmospheric Data. *Mon. Weath. Rev.* **129**:2116-2125.
- Pedley, T. J. 1997. Introduction to fluid dynamics. *Sci. Mar.* **61**:7-24.
- Petchey, O. L. 2000. Environmental colour affects aspects of single-species population dynamics. *Proc. R. Soc. Lond. B* **267**:747-754.
- Petchey, O. L., P. T. McPhearson, T. M. Casey, and P. J. Morin. 1999. Environmental warming alters food-web structure and ecosystem function. *Nature* **402**:69-72.
- Pham, D. T. 2001. Stochastic methods for sequential data assimilation in strongly non-linear systems. *Mon. Weath. Rev.* **129**:1194-1207.
- Phillips, J. B., M. E. Deutschlander, M. J. Freake, and S. C. Borland. 2001. The role of extraocular photoreceptors in newt magnetic compass orientation: parallels between light-dependent magnetoreception and polarized light detection in vertebrates. *J. Exp. Biol.* **204**:2543-2552.
- Piterbarg, L. I., and T. M. Ozgökmen. 2002. A simple prediction algorithm for the lagrangian motion in two-dimensional turbulent flows. *J. Appl. Math.* **63**:116-148.
- Planque, B., and A. H. Taylor. 1998. Long-term changes in zooplankton and the climate of the North Atlantic. *ICES J. Mar. Sci.* **55**:644-654.
- Pollard, R. T., and L. A. Regier. 1992. Vorticity and vertical circulation at an ocean front. *J. Phys. Oceanogr.* **22**:609-625.
- Polovina, J. J., D. R. Howell, D. R. Kobayashi, and M. P. Seki. 2001. The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. *Progress In Oceanography* **49**:469-483.
- Polovina, J. J., D. R. Kobayashi, D. M. Parker, M. P. Seki, and G. H. Balazs. 2000. Turtles on the edge: movement of loggerhead turtles (*Caretta caretta*) along oceanic fronts, spanning longline fishing grounds in the central North Pacific, 1997-1998. *Fisheries Oceanography* **9**:71-82.
- Popper, S. K. 1982. *La connaissance objective*. Editions Complexe.
- Press, W. H. 1978. *Comments Astrophys.* **7**:103.
- Prieur, L. 1985. Systèmes frontaux en Mer Ligure à partir des mesures multiparamétriques en continu. *Le courrier de Médiprod* **2**:2-4.
- Pyke, G. H. 1984. Optimal foraging theory: a critical review. *Ann. Rev. Ecol. Syst.* **15**:523-575.
- Rai, V., and W. M. Schaffer. 2001. Chaos in ecology. *Chaos, solitons and fractals* **12**:197-203.
- Ranta, E., P. Lundberg, V. Kaitala, and J. Laakso. 2000. Visibility of the environmental noise modulating population dynamics. *Proc. R. Soc. Lond. B* **267**:1851-1856.
- Ravier, C. 2003. Fluctuations à long terme du thon rouge - Validité, origines, conséquences. PhD. ENSAR.
- Ravier, C., and J.-M. Fromentin. 2001. Long-term fluctuations in the Eastern Atlantic and Mediterranean bluefin tuna population. *ICES J. Mar. Sci.* **58**:1299-1317.
- Ravier, C., and J.-M. Fromentin. 2004. Are the long term fluctuations in Atlantic bluefin tuna population related to environmental changes? *Fisheries Oceanography* **13**:145-160.

- Reid, G. C. 1997. Solar forcing of global climate change since the mid-17th century. *Clim. Change* **37**:391-405.
- Reynolds, R. W., and T. M. Smith. 1994. Improved global sea surface temperature analyses using optimum interpolation. *J. Climate* **7**:929-948.
- Reynolds, W. W., and M. E. Casterlin. 1976. Behavioral thermoregulation and the 'final preferendum' paradigm. *American Zoologist* **19**:211-224.
- Rigby, S., and C. V. Mislom. 2000. Origins, evolution, and diversification of zooplankton. *Annu. Rev. Ecol. Syst.* **31**:293-313.
- Rinaldi, S., and C. Solidoro. 1998. Chaos and peak-to-peak dynamics in a plankton-fish model. *Theor. Pop. Biol.* **54**:62-77.
- Ripa, J., and P. Lundberg. 1996. Noise colour and the risk of extinctions. *Proc. R. Soc. Lond. B* **263**:1751-1753.
- Ripa, J., P. Lundberg, and V. Kaitala. 1998. A General Theory of Environmental Noise in Ecological Food Webs. *Am. Nat.* **151**:256-263.
- Ripley, B. D. 1977. Modelling spatial patterns. *J. R. Stat. Soc.* **39**:172-212.
- Risser, P. G. 1995. The status of the science examining ecotones. *Bioscience* **45**:318-325.
- Roberts, C. M., and J. P. Hawkins. 1999. Extinction risk in the sea. *Tree* **14**:241-246.
- Robinson, A. R. 1983. *Eddies in marine science*. Springer, New York.
- Rose, G. A. 2004. Reconciling overfishing and climate change with stock dynamics of Atlantic cod (*Gadus morhua*) over 500 years. *Canadian Journal of Fisheries and Aquatic Science* **61**:1553-1557.
- Rosenstein, M. T., J. J. Collins, and C. J. De Luca. 1994. Reconstruction expansion as a geometry-based framework for choosing proper delay times. *Phys. D* **73**:82-98.
- Roweis, S., and Z. Ghahramani. 1999. A Unifying Review of Linear Gaussian Models. *Neural Computations* **11**:305-345.
- Royama, T. 1992. *Analytical population dynamics*. Chapman & Hall, London.
- Rubin, D. B. 1987. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. Discussion of "The calculation of posterior distributions by data augmentation" by Tanner and Wong. *Journal of the American Statistical Association* **82**:543-546.
- Saitoh, T., N. C. Stenseth, and O. Bjornstad. 1997. Density dependence in fluctuating grey-sided vole populations. *Journal of Animal Ecology* **66**:14-24.
- Santiago, J. 1998. The North Atlantic Oscillation and the recruitment of temperate tuna. *Col. Vol. Sci. Pap. ICCAT* **48**:240-249.
- Savidge, G., and H. J. Lennon. 1987. Hydrography and phytoplankton distributions in northwest Scottish waters. *Continental Shelf Res.* **7**:45-66.
- Schaefer, K. M. 2001. Reproductive biology of tunas. Pages 225-270 *in* B. A. Block and E. D. Stevens, editors. *Tuna. Physiology, ecology, and evolution*. Academic Press, San Diego.

- Schaeffer, P., C. Boone, Y. Faugère, and P. Y. Le Traon. 2002. Empirical Correction of Long-Wavelength Bias in Altimeter Data Using An Optimal Analysis Method. *J. Atmos. Oceanic Technol.* **submitted**.
- Schaffer, W. M., and M. Kot. 1991. Should we expect strange attractors behind plankton dynamics - and if so should we bother? *J. Plankton Res.* **13**:1291-1305.
- Schick, R. S., J. Goldstein, and M. Lutcavage. 2004. Bluefin tuna (*Thunnus thynnus*) distribution in relation to sea surface temperature fronts in the Gulf of Maine (1994-96). *Fish. Oceanogr.* **13**:225-239.
- Schreiber, T., and A. Schmitz. 1996. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* **77**:635-638.
- Sella, M. 1929. Migrazioni e habitat del tonno (*Thunnus thynnus*, L.) studiati col metodo degli ami, con osservazioni su l'accrescimento, sul regime delle tonnare ecc. *Memoria, R. Comitato Talassografico Italiano* **156**:511-542.
- Sibert, J. R., M. K. Musyl, and R. Brill. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis from archival tagging data. *Fisheries Oceanography* **12**:141-151.
- Sibert, J. R., and A. Nielsen. 2004. The KFtrack package. *in*.
- Siegel, D. A., B. P. Kinlan, B. Gaylord, and S. D. Gaines. 2003. Lagrangian Descriptions of Marine Larval Dispersion. *Mar. Ecol. Prog. Ser.* **Submitted**.
- Simpson, J. H., and J. R. Hunter. 1974. Fronts in the Irish Sea. *Nature* **250**:404-406.
- Sims, D. W. 2003. Tractable models for testing theories about natural strategies: foraging behaviour and habitat selection of free-ranging sharks. *J. Fish Biol.* **63**:53-73.
- Sivakumar, B. 2000. Chaos theory in hydrology: important issues and interpretations. *Journal of Hydrology* **227**:1-20.
- Skud, B. E. 1982. Dominance in fishes: the relation between environment and abundance. *Science* **200**:403-411.
- Sokal, R. R., N. L. Oden, B. A. Thomson, and J. H. Kim. 1993. Testing for regional differences in means: distinguishing inherent from spurious spatial autocorrelation by restricted randomization. *Geographical Analysis* **25**:199-210.
- Solé, R. V., and J. Bascompte. 1995. Measuring chaos from spatial information. *J. Theor. Biol.* **175**:139-147.
- Spagnolo, B., D. Valenti, and A. Fiasconaro. 2004. Noise in ecosystems: a short review. *Mathematical Biosciences and Engineering* **1**:185-211.
- Spencer, P. D., and J. S. Collie. 1997. Patterns of population variability in marine fish stocks. *Fisheries Oceanogr* **6**:188-204.
- Spiegelhalter, D. J., A. Thomas, N. Best, and W. R. Gilks. 1996. BUGS 0.5, Bayesian inference using Gibbs sampling. Manual version ii Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge.
- Steele, J. H. 1985. A comparison of terrestrial and marine ecological systems. *Nature* **313**:355-358.
- Steele, J. H. 1989. The ocean "landscape". *Landscape Ecology* **3**:185-192.

- Steele, J. H. 1995. Can ecological concepts span the land and ocean domains? Pages 5-19 *in* T. M. Powell and J. H. Steele, editors. *Ecological time series*. Chapman & Hall, New York.
- Steele, J. H., and E. W. Henderson. 1984. Modeling long-term fluctuations in fish stocks. *Science* **224**:985-987.
- Stenseth, N. C., O. Bjørnstad, W. Falck, J.-M. Fromentin, J. Gjøsæter, and J. Gray. 1999. Dynamics of coastal cod populations: intra- and inter- density-dependence and stochastic processes. *Proceedings of the Royal Society London, Serie B* **266**:1645-1654.
- Stenseth, N. C., A. Mysterud, G. Ottersen, J. W. Hurrell, K.-S. Chan, and M. Lima. 2002. Ecological Effects of Climate Fluctuations. *Science* **297**:1292-1296.
- Stenseth, N. C., G. Ottersen, J. W. Hurrell, A. Belgrado, and (eds). 2004. *Marine ecosystems and climate variations - the North Atlantic. A comparative perspective*. Oxford University Press, Oxford.
- Stromberg, J.-O. 1997. Human influence or natural perturbation in oceanic or coastal waters - can we distinguish them? *Hydrobiologia* **352**:181-193.
- Student, L. R. 1914. The elimination of spurious correlation due to position in time or space. *Biometrika* **10**:179-181.
- Takens, F. 1981. Detecting strange attractors in turbulence. Pages 366-381 *in* D. A. Rand and L.-S. Young, editors. *Dynamical Systems and Turbulence*. Springer-Verlag, New York.
- Takeoka, H., O. Matsuda, and T. Yamamoto. 1993. Processes causing the chlorophyll a maximum in the tidal front in Iyonada, Japan. *J. Oceanogr.* **49**:57-70.
- Teo, S. L. H., A. Boustany, S. B. Blackwell, A. Walli, K. C. Weng, and B. A. Block. 2004. Validation of geolocation estimates based on light level and seas urface temperature from electronic tags. *Mar. Ecol. Prog. Ser.* **283**:81-98.
- The Ring Group. 1981. Gulf Stream cold core rings: their physics, chemistry and biology. *Science* **212**:1091-1100.
- Theiler, J., and S. Eubank. 1993. Don't bleach chaotic data. *Chaos, Solitons & Fractals* **3**:771-782.
- Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. 1992. Testing for non-linearity in time series: the method of surrogate data. *Phys. D* **58**:77-94.
- Tiews, K. 1978. On the disappearance of bluefin tuna in the North Sea and its ecological implications for herring and mackerel. *Rapport et Procès-verbaux des Réunions du Conseil international de l'Exploration de la Mer* **172**:301-309.
- Torrence, C., and G. P. Compo. 1998. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**:61-78.
- Toumazou, V., and J. Cretaux. 2001. Using a Lanczos eigensolver in the computation of empirical orthogonal functions. *Mon. Weath. Rev.* **125**:1243-1250.
- Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman. 2001. Missing Value estimation methods for DNA microarrays. *Bioinformatics*. *Bioinformatics* **17**:520-525.
- Trulla, L. L., A. Giuliani, J. P. Zbilut, and C. L. Webber Jr. 1996. Recurrence quantification analysis of the logistic equation with transients. *Physics Letters A* **233**:255-260.
- Turchin, P. 1996. Fractal analysis of animal movement: a critique. *Ecology* **77**:2086-2090.

- Turchin, P. 2000. *Complex Population Dynamics: a Theoretical/Empirical Synthesis*. Princeton.
- Turchin, P. 2003. *Complex Population Dynamics: a Theoretical/Empirical Synthesis*. Princeton.
- Turchin, P., and S. P. Ellner. 2000. Living on the edge of chaos: population dynamics of Fennoscandian voles. *Ecology* **81**:3099-3116.
- Tyler, J. A., and W. W. Hargrove. 1997. Predicting spatial distribution of foragers over large resource landscapes: a modeling analysis of the Ideal Free Distribution. *Oikos* **79**:376-386.
- Ullman, D. S., and P. Cornillon. 1999. Satellite-derived sea surface temperature fronts on the continental shelf off the northeast U.S. coast. *Journal of Geophysical Research* **104**:23459-23478.
- Ullman, D. S., and P. Cornillon. 2000. Evaluation of front detection methods for satellite-derived SST data using *in situ* observations. *J. Atmos. Oceanic Technol.* **17**:1667-1675.
- van de Plassche, O., K. van der Borg, and A. F. M. de Jong. 1998. Sea level-climate correlation during the past 1400 years. *Geology* **26**:319-322.
- van de Plassche, O., G. van der Schrier, S. L. Weber, W. R. Gehrels, and A. J. Wright. 2003. Sea-level variability in the northwest Atlantic during the past 1500 years: a delayed response to solar forcing? *Geophysical Research Letters* **30**:1-4.
- van der Schrier, G., S. L. Weber, and S. S. Drijfhout. 2002. Sea level changes in the North Atlantic by solar forcing and internal variability. *Climate Dynamics* **19**:435-447.
- Vasseur, D. A., and P. Yodzis. 2004. The color of environmental noise. *Ecology* **85**:1146-1152.
- Venema, V., S. Meyer, S. Gimeno García, A. Kniffka, C. Simmer, S. Crewell, U. Löhnert, T. Trautmann, and A. Macke. submitted. Surrogate cloud fields generated with the Iterative Amplitude Adapted Fourier Transform algorithm.
- Verity, P. G., V. Smetacek, and T. J. Smayda. 2002. Status, trends and the future of the marine pelagic ecosystem. *Environmental Conservation* **29**:207-237.
- Videler, J. J., and D. Weihs. 1982. Energetic advances of burst-and-coast swimming of fish at high speeds. *J. Exp. Biol.* **97**:169-178.
- Von Bertalanffy, L. 1968. *Théorie générale des systèmes*. Dunod.
- Wagner, F. H. 1996. Principles for the conservation of wild living resources: another perspective. *Ecological Applications* **6**:365-367.
- Walker, G. 1931. On Periodicity in Series of Related Terms. *Proc. R. Soc. Lond. A* **131**:518-532.
- Walker, M. M. 1984. Learned magnetic field discrimination in the yellowfin tuna, *Thunnus albacares*. *J. Comp. Physiol. A.* **155**:673-679.
- Wang, D., Y. Liu, Y. Qi, and P. Shi. 2001. Seasonal Variability of Thermal Fronts in the Northern South China Sea from Satellite Data. *Geophysical Research Letters* **28**:3963-3966.
- Wardle, C. S., J. J. Videler, T. Arimoto, J. M. Franco, and P. He. 1989. The muscle twitch and the maximum swimming speed of giant bluefin tuna, *Thunnus thynnus*. *J. Fish Biol.* **35**:129-137.
- Webber Jr., C. L., and J. P. Zbilut. 1994. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology* **76**:965-973.

- Weihls, D. 1974. Energetic advances of burst swimming in fish. *J. Theor. Biol.* **48**:215-229.
- White, W. B., L. J., D. R. Cayan, and M. D. Dettinger. 1997. Response of global temperature to changing solar irradiance. *J. Geophys. Res.* **102**:3255-3266.
- Wiens, J. A., C. S. Crawford, and J. R. Gosz. 1985. Boundary dynamics: a conceptual framework for studying landscape ecosystems. *Oikos* **45**:412-427.
- Wiens, J. A., T. O. Crist, K. A. With, and B. T. Milne. 1993. On quantifying insect movements. *Environmental Entomology* **22**:709-715.
- Wildhaber, M. L., and L. B. Crowder. 1990. Testing a bioenergetics-based habitat choice model: bluegill (*Lepomis macrochirus*) responses to food availability and temperature. *Can. J. Fish. Aquat. Sci.* **47**:1664-1671.
- Williams, D. W., and A. M. Liebhold. 1995. Detection of delayed density-dependence: effects of autocorrelation in an exogenous factor. *Ecology* **76**:1005-1008.
- Wilson, H. B., and D. A. Rand. 1997. Reconstructing the dynamics of unobserved variables in spatially extended systems. *Proc. R. Soc. Lond. B* **264**:625-630.
- Wilson, R. P., J. J. Ducamp, W. G. Rees, B. M. Culik, and K. Niekamp. 1992. Estimation of location: global coverage using light intensity. Pages 131-135 in I. G. Priede and S. M. Swift, editors. *Wildlife Telemetry: Remote Monitoring and Tracking of Animals*. Ellis Horwood, London.
- Wilson, S. G., M. Lutcavage, R. Brill, M. P. Genovese, A. B. Cooper, and A. W. Everly. 2004. Movemens of bluefin tuna (*Thunnus thynnus*) in the northwestern Atlantic Ocean recorded by pop-up satellite archival tags. *Mar. Biol.* **in press**.
- Woodroffe, R., and J. R. Ginsberg. 1998. Edge effects and the extinction of populations inside protected areas. *Science* **280**:2126-2128.
- Young, P. C. 1998. Data-based mechanistic modeling of environmental, ecological, economic and engineering systems. *Journal of Environmental Modelling and Software* **13**:105-122.
- Yuan, X., and L. D. Talley. 1996. The subarctic frontal zone in the North Pacific: characteristic of frontal structure from climatological data and synoptic surveys. *J. Geophys. Res.* **101**:16491-16507.
- Zbilut, J. P., A. Giuliani, and C. L. J. Webber. 1998. Recurrence quantification analysis and principal components in the detection of short complex signals. *Physics Letters A* **237**:131-135.
- Zbilut, J. P., N. Thomasson, and C. L. Webber Jr. 2002. Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. *Medical Engineering & Physics* **24**:53-60.
- Zbilut, J. P., and C. L. Webber Jr. 1992. Embedding and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**:199-203.

Articles et manuscrits

1. Royer, F. **Seasonal and spatial patterns of thermal fronts in the NE Atlantic and the Mediterranean sea.** *Manuscrit*
2. Royer, F., Fromentin, J.-M., and Gaspar, P. (2004) **Association between bluefin tuna schools and oceanic features in the Gulf of Lions.** *Marine Ecology Progress Series.* 269: 249-263
3. Royer, F., Fromentin, J.-M., and Gaspar, P. (2005) **A state-space model to infer bluefin tuna movement and habitat from archival tags.** *Oikos* 109: 473-484
4. Royer, F. **Wayfinding in the ocean: sensors for the equipment of archival tags.** *Manuscrit*
5. Royer, F. and Fromentin, J.-M. **Recurrence analysis in ecological time series: an application to Atlantic bluefin tuna.** *Manuscrit*
6. Royer, F., and Fromentin, J.-M. **Environmental noise and Spawning Strategies: the case of Atlantic Bluefin Tuna.** *Manuscrit*

Autres contributions

Royer, F., Fromentin, J.-M., Farrugio, H. and Gaspar, P. (in press) **Determining bluefin tuna habitat through frontal features in the Mediterranean sea.** *Col. Vol. Sci. ICCAT*

Fromentin, J.-M., Royer, F. and Marsac, F. (in press) **Open-source environmental databases for open-sea fisheries biologists.** *Col. Vol. Sci. ICCAT*

Posters et communications orales

Royer, F., Fromentin, J.-M., Gaspar, P. and Larnicol, G. 2003 **Analysis of Atlantic Bluefin Tuna Behaviour in Relation to Oceanic Structures based on Remote-Sensing Data.** *Présentation orale.* ICES SYMPOSIUM ON FISH BEHAVIOUR IN EXPLOITED ECOSYSTEMS - BERGEN

Royer, F., Farrugio, H., Fromentin, J.-M. and Gaspar, P. 2004 **Le Thon Rouge de Méditerranée: Association avec les Systèmes Frontaux du Panache Rhodanien.** *Poster.* COLLOQUE RHONE-MEDITERRANEE - MARSEILLE

Royer, F., Fromentin, J.-M. and Gaspar, P. 2004 **Modelling bluefin tuna behaviour in its environment: can a generalized state-space model resolve estimation problems in archival tagging?** *Présentation orale.* 55th TUNA CONFERENCE – LAKE ARROWHEAD, CA

Royer, F., Farrugio, H., Fromentin, J.-M. and Gaspar, P. 2004 **Association between juvenile bluefin tuna schools and frontal features in the Gulf of Lions.** *Poster.* 55th TUNA CONFERENCE – LAKE ARROWHEAD, CA

Royer, F., Gaspar, P., Fromentin, J.-M. and Lutcavage, M. 2004 **Modélisation de la dynamique spatiale du thon rouge Nord-Atlantique par couplage de marques archives et de sorties Mercator.** *Poster.* REUNION DU GROUPE MISSION MERCATOR – METEO FRANCE, TOULOUSE

Royer, F. and Farrugio, H. 2004 **Sur les traces d'un nomade des mers, le thon rouge.** *Poster de vulgarisation scientifique.* FETE DE LA SCIENCE – UNIVERSITE DE MONTPELLIER

**Seasonal and Spatial Patterns of Thermal Fronts
in the North-Eastern Atlantic and the Mediterranean Sea**

Manuscript

F. Royer^{1,2}

¹IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean Monnet, BP 171, 34203 Sète cedex, France

²CLS, Division Océanographie Spatiale, 8-10 rue Hermès, 31526 Ramonville St Agne, France

1 **Abstract**

2

3 The spatial and temporal variability of thermal fronts in the North-Eastern Atlantic and the
4 Mediterranean Sea is investigated in this study: a data set of daily SST maps with a ground
5 resolution of 2km/pixel is first constructed using Local Area Coverage products. A state-of-
6 the-art edge detection algorithm involving histogram and cohesion analysis is then applied to
7 each image. A quarterly climatology of thermal fronts is computed, covering the 2001-2004
8 period. The seasonal and spatial patterns of fronts are then investigated using a decomposition
9 in orthogonal functions. Four regions, centered on the bay of Biscaye, the Canarian
10 archipelago and covering the western and eastern Mediterranean basins are compared. All
11 basins display a complex array of frontal patterns: a strong seasonal component is highlighted,
12 in association with coastal upwellings and orographic windy events. This heterogeneous
13 pattern is complemented by the other EOF modes, which appear to characterize more
14 ubiquitous fronts and jets, associated with transient mesoscale eddies and events of surface
15 circulation. The seasonal component is in particular extracted by means of EOFs, and shown
16 to correlate to narrow shelf-break regions or windy areas. The interannual component clearly
17 differentiates each year from another, and is likely to be related by large-scale changes in
18 surface wind stress, possibly in relation to the North Atlantic Oscillation.

19

20 **Index terms**

21 Fronts and jets, instruments and techniques, Ecosystems, structure, dynamics and modeling

1

1 **Introduction**

2 Surface fronts are ubiquitous features in the ocean, where enhanced mixing and air-sea
3 exchanges can occur, as well as higher primary productions and a greater activity of upper
4 trophic levels. A wealth of studies in the 80-90s reported the biophysical characteristics or the
5 cross-structure of oceanic fronts during at-sea surveys, while more recent papers deal with
6 their regional distribution and seasonality, as derived from satellite-based radiometers. Their
7 identification and delineation has greatly benefited from the availability of such data and the
8 advent of new edge detection algorithms. As such, a front is not only a region of high
9 geophysical gradients but also the interface between two neighbouring and distinct water
10 masses. Earlier algorithms employed gradient operators which are sensitive to noise and relied
11 on user defined thresholds [*Holyer and Peckinpaugh, 1989; Ullman and Cornillon, 2000*].
12 State-of-the-art methods [*Cayula and Cornillon, 1992; Cayula and Cornillon, 1995; Shaw
13 and Vennell, 2000*] explicitly define fronts as interfaces, allowing for a more objective
14 mapping of these features at local scales (Cayula et al., 1991). Recent achievements included
15 the description of frontal features in shelf or marginal seas [*Belkin and Cornillon, 2003;
16 Belkin et al., 2004; Ullman and Cornillon, 1999*] and the correlation with atmospheric or
17 oceanic forcing sources [*Ullman and Cornillon, 1999; Wang et al., 2001*]. Further
18 applications of edge mapping are: i) the assimilation of frontal location data for improved
19 modelling of non-linear circulation patterns (Le Traon et al.), and ii) the development of a
20 “feature-oriented” description of habitat in marine ecosystems studies [*Arrizabalaga et al.,
21 2002; Etnoyer et al., 2004; Royer et al., 2004; Schick et al., 2004*].

22

23 This work aims at comparing the spatial structure and seasonal cycle of thermal fronts
24 in the Mediterranean sea and in the North-Eastern Atlantic Ocean (in a region ranging from
25 Iceland to the Canarian archipelago). These two zones are subject to atmospheric forcing of
26 very different amplitude and scales, with a wealth of hydrographic and bathymetric options,
27 and are therefore expected to display diverse modes of frontal variability. A daily data set of
28 high resolution infrared images were processed, yielding a monthly overview of thermal
29 fronts density between june, 2001 and august, 2004. Four regions were considered, covering
30 the area under study. Annual time series were first derived from this data set to assess the
31 seasonal cycles. To further investigate the interannual variability of front density, we
32 employed the sea-level pressure difference between Gibraltar and Iceland, thus hypothesizing a

1 regional effect of the North Atlantic Oscillation on the homogeneity of surface waters. A
2 decomposition in Empirical Orthogonal Functions was then performed to assess the
3 space/time structure of surface fronts in all four areas.

4 5 **Data**

6 Sea Surface Temperature maps are routinely provided by the Ocean & Sea Ice Satellite
7 Application Facility, using data collected by the AVHRR sensor on board of NOAA 14, 16
8 and 17. The Centre de Météorologie Spatiale of Lannion (France) distributes LAC Level-3
9 products (Local Area Coverage in NCSA's Hierarchical Data Format) on a daily basis at a
10 ground resolution of 2 km, for the six following areas: 1) MORI – Eastern Mediterranean, 2)
11 MOCC – Western Mediterranean, 3) CANA – North-East Atlantic from Portugal down to the
12 Canarian archipelago, 4) GASC – North-East Atlantic from Portugal to Iceland, 5) NORV –
13 North-East Atlantic between Greenland and Norway, and 6) MNOR – North Sea and Baltic
14 Sea. Radiance measures from AVHRR's channels 3, 4 and 5 are processed using the protocol
15 defined by Brisson et al. (1981), resulting in a night bias of +/- 0.4K [Brisson et al., 2001].
16 Cloud masking is performed using a fine scale climatology of surface temperature minima
17 and a spatial gradient climatology. All daily images from June, 2001 to July, 2004 were
18 collected for this study, regardless of cloud coverage.

19 20 **Methods**

21 Daily SST maps were further processed using Cayula and Cornillon's Single Image
22 Edge Detection technique, [Cayula and Cornillon, 1992] which operates using overlapping
23 windows: neighbouring water masses are identified by i) testing for histogram bimodality, and
24 ii) ensuring the cohesiveness of the two pixel populations. As an improvement, window sizes
25 can be made variable and estimated locally using autocovariance or semi-variogram analysis
26 [Diehl and Wells Budd, 2000]. The 36 monthly maps that were binned by climatological
27 month and year. Frontal density was computed as a probability of occurrence for each pixel,
28 *i.e.* as the number of times a pixel was classified as a front divided by the number of times this
29 pixel was cloud-free [Belkin et al., 2004]. Note that in relatively cloud-free zones (*e.g.* the
30 Mediterranean basins), gradient-based and histogram-based approaches are likely to have
31 converging results when binning data over large time periods (*e.g.* in monthly maps).
32 However, this is not the case for regions hampered by nebulosity: spurious phenomena, if
33 undetected (*e.g.* unmasked noisy pixels, cloud borders) are likely to generate a significant

1 number of false detections when relying solely on gradient thresholding, leading to over-
2 estimation of monthly/annual frontal density. An histogram-based algorithm appears more
3 appropriate for frontal climatologies in cloudy zones (e.g. the north and Baltic seas, the NE
4 Atlantic and the bay of Biscay), rather than considering spatial gradients over monthly periods
5 [*Etnoyer et al.*, 2004; *Wang et al.*, 2001].

6 To extract the main frontal patterns in each zone, Empirical Orthogonal Functions
7 (EOFs) were applied to the daily dataset binned by month. EOFs are widely used in the
8 analysis of geophysical variables, and allow to decompose statistical variability over
9 measurements and samples using orthogonal series (Lawson and Hanson, 1974). The Lanczos
10 method was used for this purpose [*Toumazou and Cretaux*, 2001]. Fronts and interfaces are
11 typically transient features, and monthly averaged (smoothed) maps are expected to
12 characterize only the most stable fronts: despite their highly non-linear nature and multiple
13 origins, an EOF decomposition appears suitable for the identification of both seasonal and
14 transient components. Front density fields are also characterized by large variances, since
15 edge pixels are (from a statistical viewpoint) as variant as a sum of Dirac delta functions.
16 Monthly maps were therefore binned to a coarser ground resolution of 4x4 km, thus
17 stabilizing the variance and speeding up computations.

18

19 **Results**

20 For general comparison purposes, monthly maps were aggregated over two semesters
21 spanning May to October, and November to April. This provides an overview of the frontal
22 activity in the studied area: high densities are clearly visible at the south-eastern boundaries of
23 the two Atlantic regions, forming the most intense pattern of the dataset. At the other
24 extremity of our window, high local densities draw the attention to the north-east of Crete.
25 The Mediterranean sea appears somewhat homogeneous with both ubiquitous fronts and wide
26 patches covering the two basins. To the North-West, noisy pixel clouds corresponding to little
27 or no detections cover a large part of the window. Careful examination allows to distinguish a
28 number of secondary patterns over the area, yet faded by the common colour scale employed
29 for all regions.

30 The mean frontal density of each region was computed by month and by year, yielding
31 eight annual time series: a seasonal cycle can be readily identified, with a low frontal activity
32 during the summer months, and high densities during winter. Values appear generally the
33 highest for the Bay of Biscaye and North-East Atlantic region, since this area is more likely to

1 be covered by clouds, thus biasing frontal estimates. Still, a striking feature emerges with the
2 classification of each year: the four regions display a common interannual pattern, with the
3 year 2001 showing the strongest signal, with a high amplitude and peaks, along with the year
4 2002. The years 2003 and 2004 display a much weaker and relatively homogeneous signal.
5 Despite the incompleteness of the years 2001 and 2004 in our study, these patterns are clearly
6 distinguishable and shared by all four data sets. Since these four regions cover a wide area, a
7 highly plausible teleconnection explaining linked dynamics at these scales would involve
8 atmospheric forcing. As such, a link with the major climatic phenomena in the northern
9 hemisphere, *i.e.* the NAO, was hypothesized. To account for NAO fluctuations over our study
10 area, the difference in sea level pressure between Iceland and Gibraltar was computed, rather
11 than the classic Azores-based index. As monthly NAO indices are usually highly noisy, a 6-
12 month running was applied to the time series, thus allowing for easier interpretations. It
13 appeared that the years 2001 and 2002 are characterized by positive or weakly negative NAO
14 values, while the years 2003 and 2004 show negative or weakly positive NAO values. While
15 this correlation does not imply causation, an indirect effect of atmospheric fluxes on the
16 stability of surface water masses can be postulated.

17 However, frontogenesis in the four regions of this study may originate from a great
18 variety of physical mechanisms (*e.g.* NE wind-driven coastal upwellings in the Canarian area,
19 NW offshore winds interacting with the Liguro-Provencal current in the Western
20 Mediterranean, or shelf-sea fronts affected by onshore west winds in the bay of Biscaye).
21 Several different mechanisms, each responding negatively or positiveley to wind stress and
22 spatially varying, are therefore involved. In an attempt to better characterize the space/time
23 distribution of surface fronts, EOFs were therefore applied to each region of the study: results
24 are presented in Fig. ? and ?. The four leading spatial modes are presented, along with their
25 temporal component and the percent of variance explained by each axis. The corresponding
26 space and time patterns is further detailed.

27 **The Canarian Region.** The first EOF mode yields 37% of the total variance, and
28 corresponds to well-delimited patterns of wide frontal systems running along the coasts, and
29 smaller patches to the South-West of each island of the Azores and Canarian archipelago.
30 This mode is associated to a strong seasonal cycle, with sinusoidal fluctuations peaking
31 around June-August. The second EOF mode accounts for 23% of the total variance, and is
32 much more spatially homogeneous than the first one. Still, coherent structures can be seen in
33 open ocean areas, especially in the northern part of the window, forming a strip departing
34 from the Portugal coast. Almost no seasonality appears in this temporal signature: the same

1 comment can be made about the third spatial mode, which also displays discernable structures
2 in open ocean regions. The fourth mode is the noisiest of all, and does not display any
3 cyclicity either.

4 **The NA Drift Region and the Bay of Biscaye.** This region is characterized by wider
5 shelf areas and protected shelf seas, and is therefore more subject to tidal mixing. A greater
6 complexity in the patterns of frontal density can be seen in the four EOFs modes. In
7 particular, shelf-break fronts are highly visible in the Bay of Biscaye, off the coast of Brittany
8 and to the North-West of Ireland. An array of meso-scale features can also be detected off-
9 shore, in both the second and third spatial EOFs. Coastal fronts are revealed by all spatial
10 maps, and are not clearly disentangled by the decomposition. In the same time, the amount of
11 variance explained by each axis is relatively constant (*i.e.* from 32 to 22%). While the first
12 mode characterized the same seasonal cycle as above, with peaks occurring in June-August,
13 the other modes are also marked by weak pseudo-cycles.

14 **The Western Mediterranean basin.** A well-defined pattern arises from the first EOF
15 spatial mode, accounting for 55% of the total variance: marked frontal densities are visible in
16 the Alboran sea, running close to the Algerian coast and south to Sicily. High densities
17 surround Sardinia and Corsica, with an isolated hotspot east to Bonifacio. The path bordering
18 the Ligurian-Provencal current then draws most of the frontal intensities, as far as the Catalan
19 coast. This first mode is associated to a regular seasonal cycle, which is marked by a steep
20 build-up phase, a wide peak in June-August, a slight increase again in October and finally a
21 smooth decreasing. A weak seasonality, with peaks occurring in October-December,
22 characterizes the second and third temporal EOFs. The corresponding spatial patterns appear
23 quite similar, and may be described jointly: frontal patches are more narrower and correspond
24 either the main axis of the Liguran-Provencal or to the delineation of the previously noted
25 "hotspot" off the Lavezzi islands. A more extended of frontal signal can also be seen to the
26 North-East of the Balearic islands. These two axis account for 18-19% of the total variance
27 each. Lastly, no clear pattern can be discerned in the fourth spatial EOF mode.

28 **The Eastern Mediterranean basin.** This region also displays wide zones of high
29 frontal densities, yet they are both less well delimited and cover greater areas. The temporal
30 decomposition effectively extracts a seasonal component in the first mode, but the cycles are
31 less well defined than in the other regions. Coastal fronts are revealed by the first spatial EOF,
32 with apparent peaks in October-January. An area of intense front densities is uncovered at the
33 eastern end of the basin. This mode accounts for 42% of the total variance. The other modes
34 account for about 20% of variance each, but they are spatially more homogeneous and noisy.

1

2 **Discussion and perspectives**

3 Ocean fronts constitute a challenging area of research, both in term of observation and
4 modelling: we have tried in this study to decompose and synthesize the year-to-year and
5 seasonal variability of thermal fronts, using mapping and decomposition in EOFs as
6 visualisation tools. The main results are that: i) a strong interannual variability signal is shared
7 by all four areas, with a clear ordering of the 2001-2002-2003-2004 years, ii) a weak seasonal
8 pattern in global frontal density, with peaks occuring during the months of june-august, iii)
9 when decomposed in space/time components, this seasonal signal accounts systematically for
10 the greatest amount of variance in the EOF decomposition.

11 The weak seasonal component visible at first in the regional time series can be
12 attributed the binning of all pixels when computing the index of each month. As fronts are
13 highly spatially heterogeneous, large, stable water masses with weak frontal densities are
14 likely to smooth out sharp signals occuring in other delimited areas. This feature is confirmed
15 and further revealed by a space/time decomposition in empirical orthogonal functions. The
16 first axis in each regions reveals how seasonal frontal areas are almost always confined to
17 narrow regions, such as upwelling areas in the Canarian region or shelf-break areas in the
18 Mediterranean basins. While this axis is evidently seasonal, the other components appear to
19 drive a mix of several spatial processes, with less clear temporal patterns. This is particularly
20 true for the North-East Atlantic region, where strong tidal effects occur: while they are likely
21 to originate from very similar processes, shelf-sea fronts appear to be separated within and
22 between axes, thus warranting any clear interpretation.

23 Higher thermal front densities are to be expected in the summer months: this can be
24 related to a decrease of regional wind stress which tends otherwise to mix surface waters and
25 blurr frontal signature. Winter is therefore the season where less densities are observed. This
26 seasonal phenomena may explain the interannual pattern in our data set: 2001 and 2002 are
27 indeed characterized by a weakly negative or positive NAO, *i.e.* a weakening in eastward
28 atmospheric flux over the Mediterranean basin, while 2002 and 2004 are characterized by a
29 weakly positive or negative NAO, *i.e.* a reinforcement of the winds over southern Europe.
30 These two last years may have therefore experienced a higher wind stress over all areas, thus
31 enhancing superficial mixing and decreasing the observed density of surface fronts. As this
32 pattern is shared by all four regions, such a regional climate phenomena is likely to be an
33 explaining factor for this distributed effect.

1 Our study covered a wide geographical area, with a great diversity in potential frontogenesis
2 mechanisms. EOFs allowed us to better extract the seasonal component and link it to specific
3 regions, but failed to reveal clear secondary patterns: more focused analysis are thus needed,
4 with the aim of isolating homogeneous frontal areas where common mechanisms are thought
5 to occur (*e.g.* shelf-sea fronts). Moreover, some frontogenic process are likely to respond
6 either positively or negatively to wind stress: *e.g.* upwelling fronts will be more visible during
7 offshore wind events (such as on the Moroccan coasts or in the Gulf of Lions), while open-sea
8 geostrophic or tidal fronts will be destroyed and regularized by any type of wind event.
9 Therefore, separating the space/time origins of thermal fronts by statistical means is likely to
10 be a difficult, data-demanding task requiring further research effort. Objective analysis
11 methods allowing to retrieve filaments and fronts in the mapping of geophysical variables are
12 strongly needed [*Alvera-Azcarate et al.*, 2004; *Beckers and Rixen*, 2003]

13

14 **Acknowledgements**

15 We thank H. Roquet and P. Le Borgne for making the high resolution dataset available from
16 the Centre de Météorologie Spatiale of Lannion (France). This study was conducted as part of
17 a Doctoral thesis thanks to a grant awarded by CLS and Ifremer to FR.

18

19

1 **Reference list**

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

- Alvera-Azcarate, A., A. Barth, M. Rixen, and J.M. Beckers, Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic Sea surface temperature, *Ocean Modelling*, *in press*, 2004.
- Arrizabalaga, H., I. Sankristobal, X. Moreno-Ventas, G. Moreno, E. Rodriguez-Martin, L.T. Kell, C. Rodriguez-Cabello, and M. Ortiz, Preliminary study on the environmental effects on bluefin catch rates in the spanish baitboat fishery of the bay of Biscay, *Col. Vol. Sci. Pap. ICCAT*, *54* (2), 620-640, 2002.
- Beckers, J.M., and M. Rixen, EOF calculations and data filling from incomplete oceanographic datasets, *J. Atmos. Oceanic Technol.*, *20*, 1839-1856, 2003.
- Belkin, I., and P. Cornillon, SST fronts of the pacific coastal and marginal seas, *Pacific Oceanography*, *1* (2), 90-113, 2003.
- Belkin, I., Z. Shan, and P. Cornillon, Global pattern of ocean fronts from Pathfinder SST data: Pacific Ocean, *In preparation*, 2004.
- Brisson, A., P. Le Borgne, and A. Marsouin, North Atlantic Regional Sea Surface Temperature Product Manual Version 1.1, pp. 30, Météo-France, Lannion, 2001.
- Cayula, J.F., and P. Cornillon, Edge detection Algorithm for SST images, *J. Atmos. Oceanic Technol.*, *9*, 67-80, 1992.
- Cayula, J.F., and P. Cornillon, Multi-image edge detection for SST images, *J. Atmos. Oceanic Technol.*, *12*, 821-829, 1995.
- Diehl, S.F., and J. Wells Budd, Geographic window sizes applied to remote-sensing sea-surface temperature front detection, pp. 1-19, Michigan Technological University, Houghton, 2000.
- Etnoyer, P., D. Canny, B. Mate, and L. Morgan, Persistent Pelagic Habitats in the Baja California to Bering Sea Ecoregion, *Oceanogr.*, *17* (1), 90-101, 2004.
- Holyer, R.J., and S.H. Peckinpaugh, Edge detection applied to satellite imagery of the oceans, *IEE Trans. Geosci. and Remote-Sens.*, *27*, 46-56, 1989.
- Royer, F., J.-M. Fromentin, and P. Gaspar, Association between bluefin tuna schools and oceanic features in the western Mediterranean, *Mar. Ecol. Prog. Ser.*, *269*, 249-263, 2004.

- 1 Schick, R.S., J. Goldstein, and M. Lutcavage, Bluefin tuna (*Thunnus thynnus*) distribution in
2 relation to sea surface temperature fronts in the Gulf of Maine (1994-96), *Fish.*
3 *Oceanogr.*, 13 (4), 225-239, 2004.
- 4 Shaw, A.G.P., and R. Vennell, A front-following algorithm for AVHRR SST imagery,
5 *Remote Sens. Environ.*, 72, 317-327, 2000.
- 6 Toumazou, V., and J. Cretaux, Using a Lanczos eigensolver in the computation of empirical
7 orthogonal functions, *Mon. Weath. Rev.*, 125 (5), 1243-1250, 2001.
- 8 Ullman, D.S., and P. Cornillon, Satellite-derived sea surface temperature fronts on the
9 continental shelf off the northeast U.S. coast, *Journal of Geophysical Research*, 104
10 (10), 23459-23478, 1999.
- 11 Ullman, D.S., and P. Cornillon, Evaluation of front detection methods for satellite-derived
12 SST data using *in situ* observations, *J. Atmos. Oceanic Technol.*, 17 (12), 1667-1675,
13 2000.
- 14 Wang, D., Y. Liu, Y. Qi, and P. Shi, Seasonal Variability of Thermal Fronts in the Northern
15 South China Sea from Satellite Data, *Geophysical Research Letters*, 28 (20), 3963-
16 3966, 2001.

17

18

19

1 **Figure captions**

2

3 Figure 1. Quarterly climatological maps of frontal densities for A. the Bay of Biscaye and
4 North-East Atlantic region, and B. the Canarian region, computed over 2001-2004: a) from
5 january to march, b) from april to june, c) from july to september, d) from october to
6 december.

7

8 Figure 2. Quarterly climatological maps of frontal densities for A. the Western Mediterranean
9 basin, and B. the Eastern Mediterranean basin, computed over 2001-2004: a) from january to
10 march, b) from april to june, c) from july to september, d) from october to december.

11

12 Figure 3. Interannual variability of thermal front density, detailed by month for a) the bay of
13 Biscaye and North-East Atlantic region, b) the Canarian region, c) the western Mediterranean
14 basin, and d) the eastern basin.

15

16 Figure 4. Monthly NAO index from january 2000 to december 2004, computed as the
17 difference in sea level pressure between Gibraltar and Iceland.

18

19 Figure 5. Decomposition in Empirical Orthogonal Functions of the monthly dataset in A. the
20 Bay of Biscaye and North-East Atlantic region, and B. the Canarian region, with: a) the four
21 leading spatial modes, b) the corresponding temporal modes, and c) the amount of variance
22 explained by each axis.

23

24 Figure 6. Decomposition in Empirical Orthogonal Functions of the monthly dataset in A. the
25 Western Mediterranean basin, and B. the Eastern Mediterranean basin, with: a) the four
26 leading spatial modes, b) the corresponding temporal modes, and c) the amount of variance
27 explained by each axis.

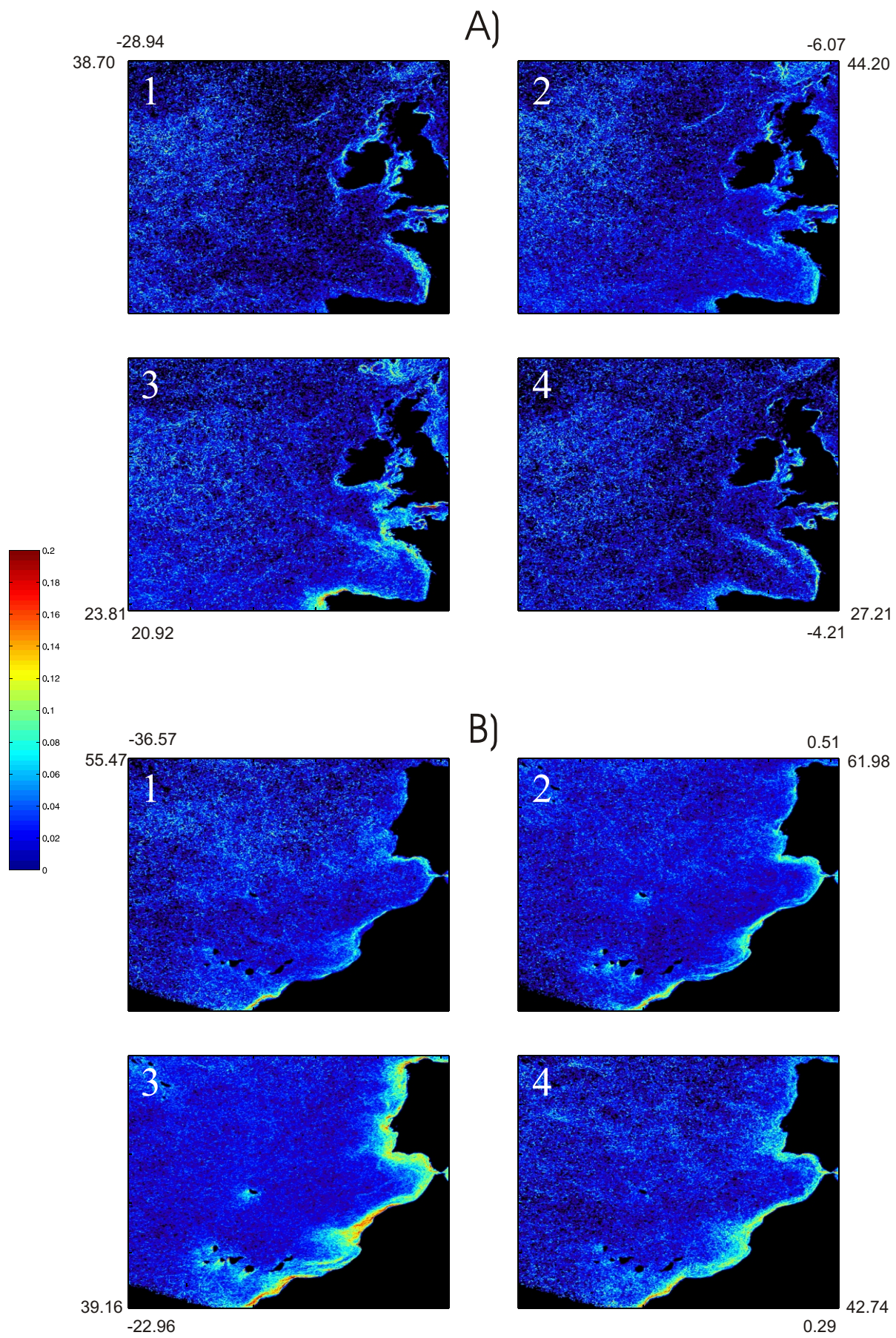
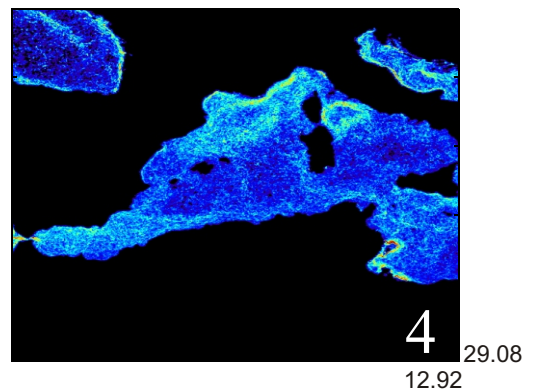
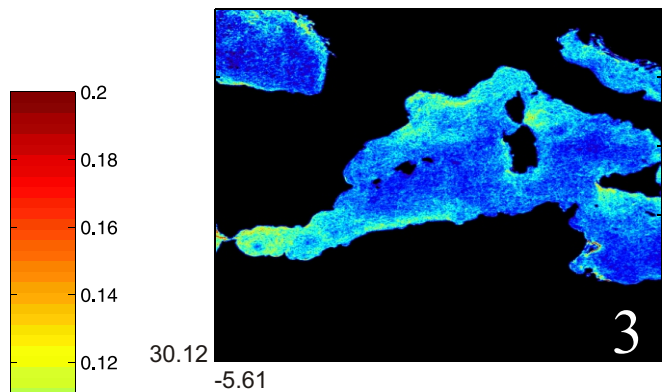
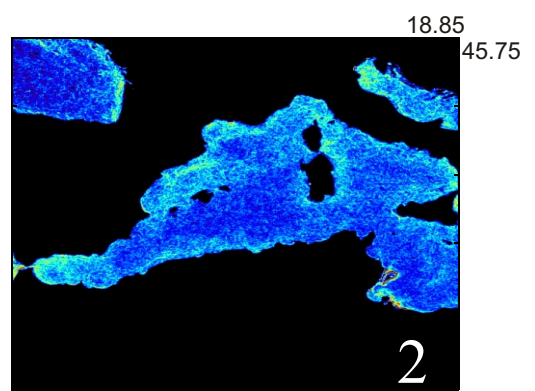
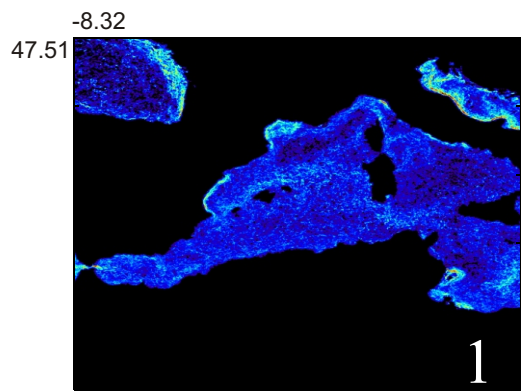


Figure 1

A)



B)

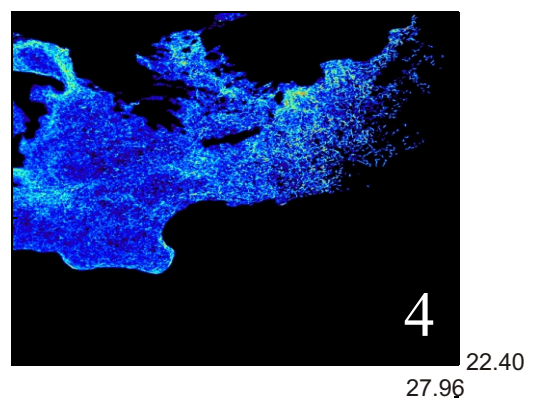
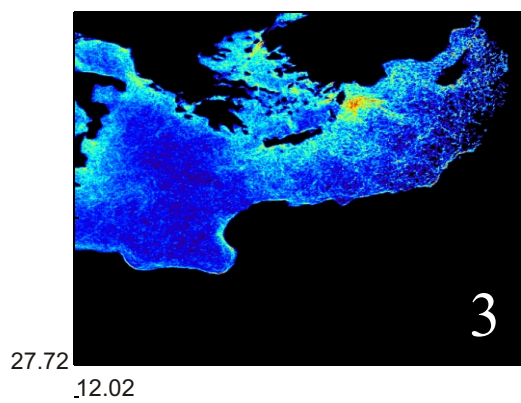
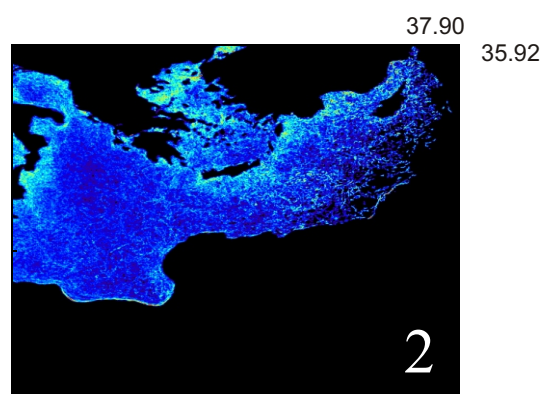
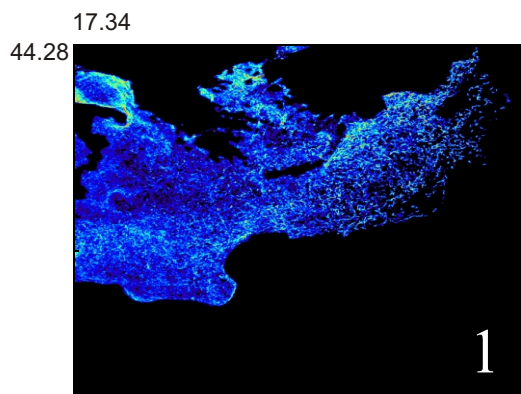


Figure 2

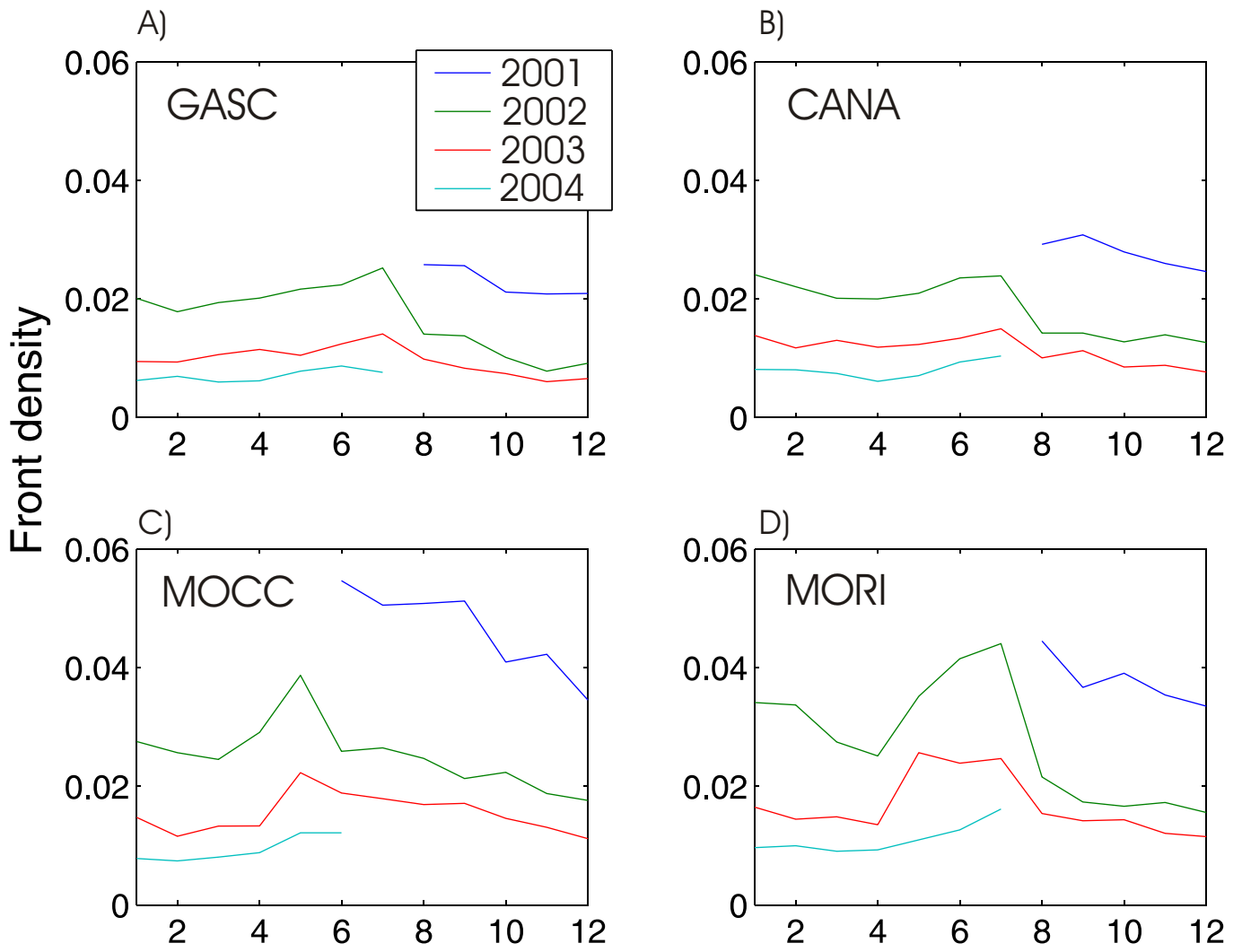


Figure 3

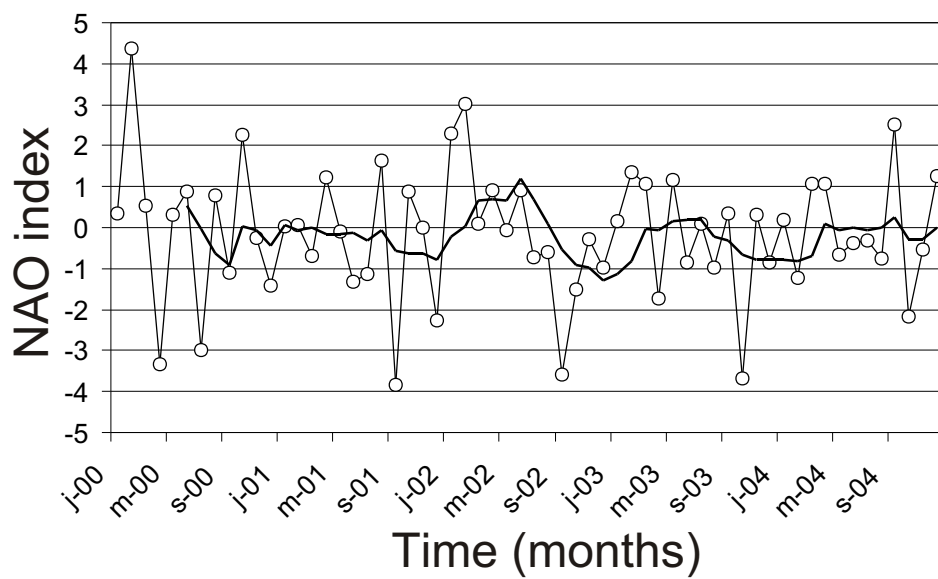


Figure 4

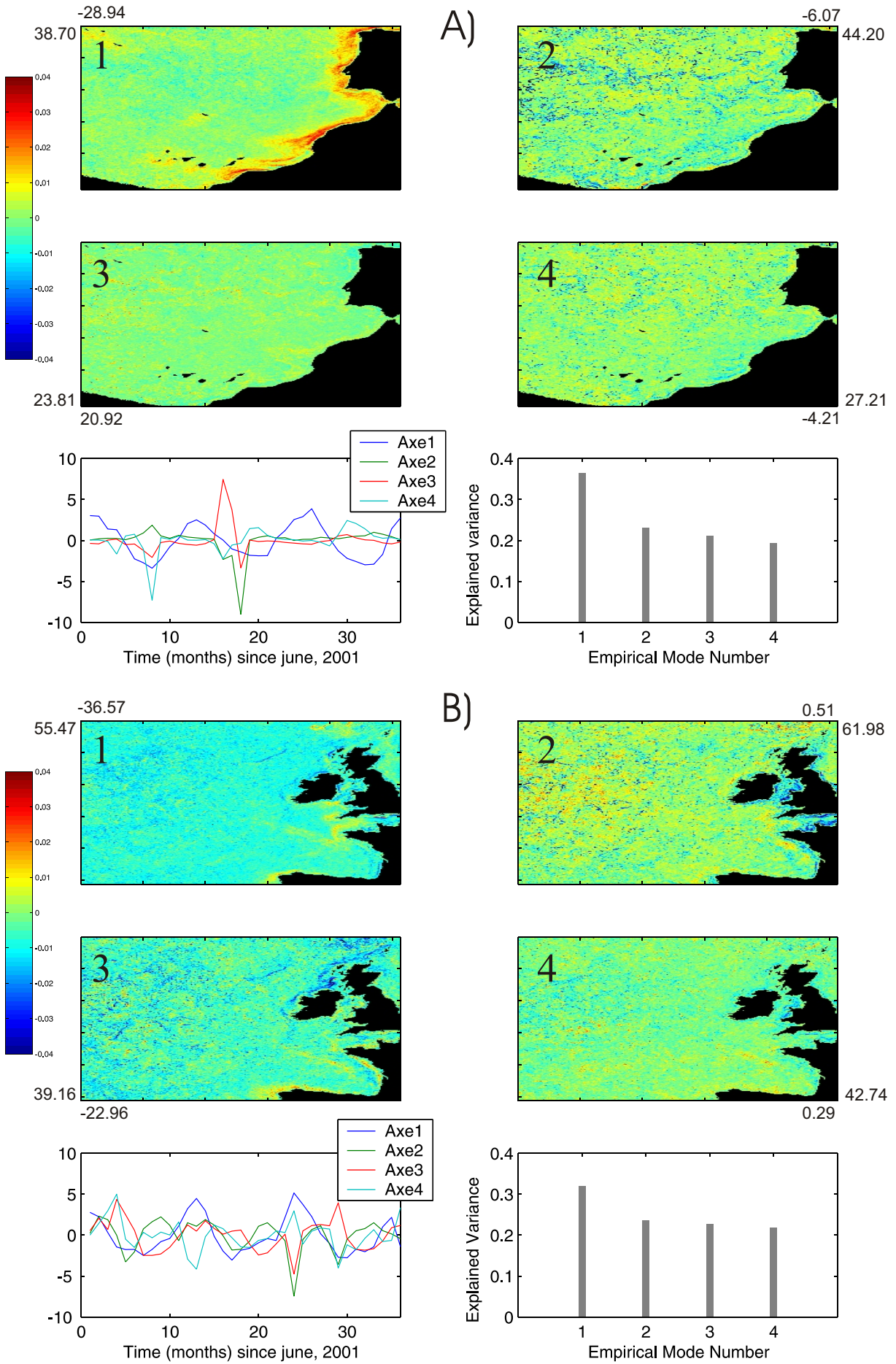


Figure 5

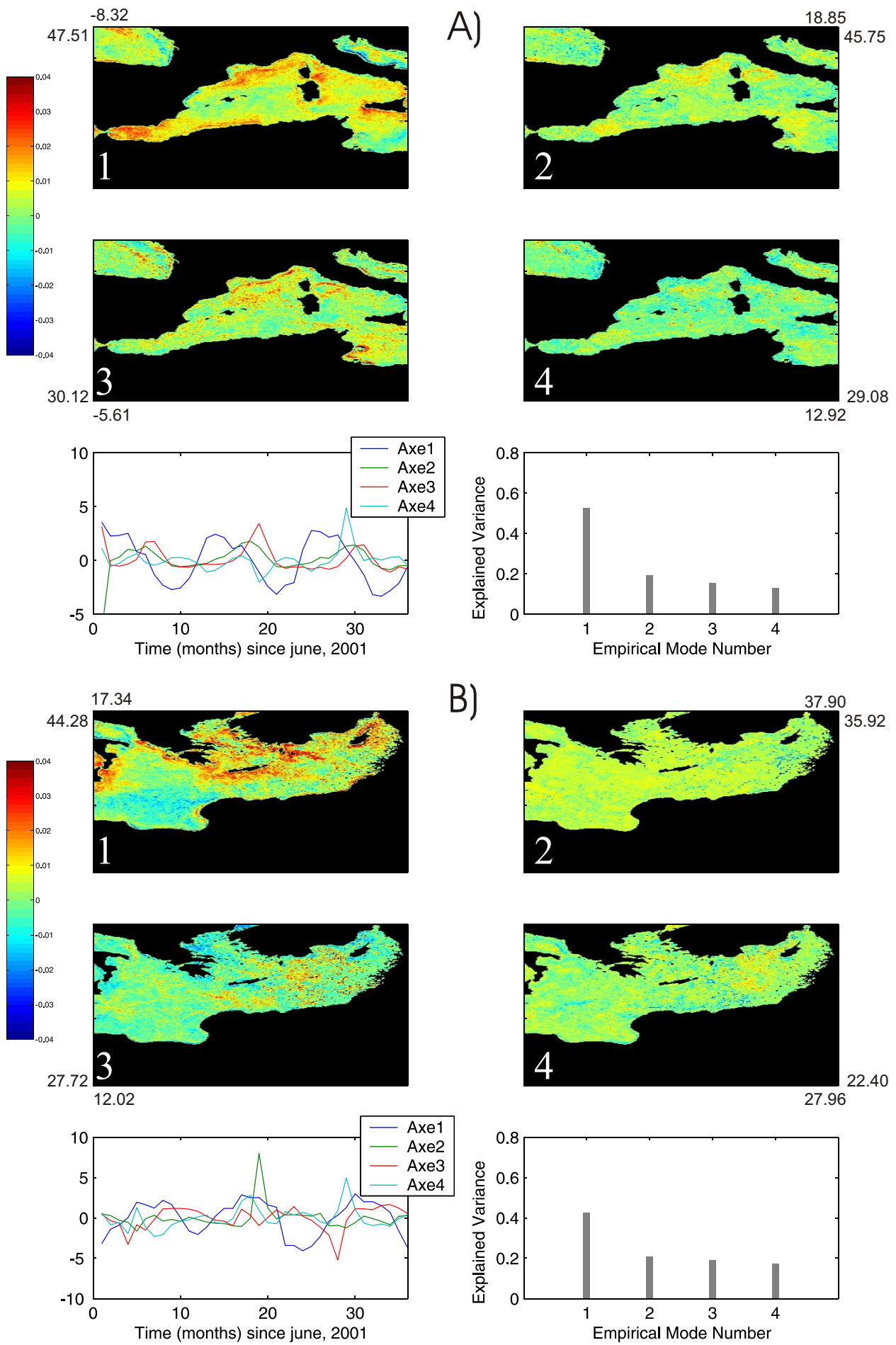


Figure 6

**ASSOCIATION BETWEEN BLUEFIN TUNA SCHOOLS
AND OCEANIC FEATURES IN THE WESTERN
MEDITERRANEAN**

F. Royer^{1,2}, J.-M. Fromentin², P. Gaspar¹

¹CLS, Division Océanographie Spatiale, 8-10 rue Hermès, 31526 Ramonville St Agne, France

²IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean
Monnet, BP 171, 34203 Sète cedex, France

Abstract

We present an analysis of the distribution of bluefin tuna (BFT) schools spotted during aerial surveys in the Gulf of Lions, in relation to oceanographic features. Bio-optical and thermal properties of the sea surface derived from high-resolution sensors (AVHRR and SeaWiFS) were studied on a daily basis, and an edge detection technique was applied to detect frontal zones. Geostatistics and point process analyses were used to evaluate the role of the environment in structuring the BFT spatial pattern. The distribution of spotted schools was strongly non-stationary both in space and time: this is believed to be an effect of the survey design (transect sampling) and the influence of transient oceanographic structures (surface fronts and eddies). The empirical variograms indicated a spatial range of the BFT schools at around 40 km, with substantial daily variability. The Ripley's K statistic, as well as the autocorrelation plots, revealed that the fish schools were clustered over a wider range of scales (from 10 to 80 km), indicating more spatial structure than expected from a random process. Finally, BFT schools distributions appeared well determined by the oceanic features, except at very small scales ($< 10\text{km}$) where over-aggregation occurred, and at the largest scales of our study ($> 40\text{km}$) where over-spreading was detected. Dynamical ecological processes, such as foraging, are likely to induce this complex spatial pattern. Possible reasons for the association of tuna with fronts are presented.

Keywords: *thunnus thynnus*, Gulf of Lions, spatial distribution, geostatistics, point process analysis, remote sensing, front detection, sea surface temperature, ocean colour, aggregation

INTRODUCTION

Atlantic bluefin tuna (BFT) is a large pelagic (top predator) and highly migratory fish living in temperate waters, usually from 35°N to 60°N (Mather et al. 1995). BFT has been continually exploited in the Mediterranean Sea since the Antiquity (Ravier & Fromentin 2001). However this species, which is relatively fragile to exploitation (Fromentin & Fonteneau 2001), is now suspected to be overfished and overexploited (ICCAT 2003), so that fishery-independent data becomes critical. Getting trustworthy information from scientific surveys, however, implies a better understanding in the relationship between BFT population dynamics (especially its spatial distribution and dynamics) and main oceanic features (Fromentin 2003). The spatial association between schools of large pelagic fish and oceanic features has been already advocated in the literature (e.g., Olson et al. 1994). For instance, surface fronts recently appeared to play a key role on the feeding, growth and physiology of large pelagic fish (e.g. Dagorn et al. 2000, Lutcavage et al. 2000). However, this type of investigation remains scarce because of a lack of adequate and precise data on the spatial distribution of fish and coincident oceanographic measurements (Laurs et al. 1984) as well as little information on the vertical dimension (our view of the oceans through remote-sensing being mainly limited to 2-D, see e.g., Brill et al. 1999, 2002). Recent advances in electronic tagging and improved 3-dimensional mapping of the ocean allowed to circumvent this last difficulty, but an in-depth analysis of the spatial distribution of BFT schools in relation to key environmental factors is still needed.

Our purpose is to test whether BFT schools distribution in the Gulf of Lions can be solely explained by key environmental variables or not. To do so, we confronted count data of Atlantic bluefin tuna obtained by aerial spottings with sea surface temperature or ocean colour obtained by high-resolution sensors (the Advanced Very High Resolution Radiometer, AVHRR and the Sea-viewing Wide Field of View Sensor, SEAWIFS). Several quantitative

methods (some being rarely applied, if ever, in marine ecology) have been used to analyse these data. An edge detection technique was developed to enhance frontal variability in the studied area. Estimates of the first and second moments (i.e. the mean and the variance) of the school distributions were also computed and modelled in relation to oceanic variability. A non-parametric measure, based on the point process framework (Diggle 1983), was applied to test possible relationships between the spatial distribution of the BFT schools (e.g. point-like data) and oceanographic fields and features (considered as continuous).

DATA

Bluefin count data from aerial surveys

As part of the EU STROMBOLI project (EU-DG XIV contract 99/022), the French Institute for the Exploitation of the Sea (IFREMER) conducted aerial surveys in the Gulf of Lions, which is a steady feeding area for juveniles (Farrugio 1981). Surveys were carried out between 2000 and 2002, in an attempt to derive an abundance index for juvenile bluefin tuna (Fromentin 2002). Such a technique appears appropriate since BFT spent most of their time in the 10 first meters (Lutcavage et al. 2000). Aerial spotting is regularly carried out since 1993 along the Southern Australian coasts on juveniles of Southern bluefin tuna (Cowling & O'Reilly 1999) and has been also tested on large Atlantic bluefin tuna of the West Atlantic (Lutcavage and Kraus 1995). Flights were performed during calm days of August and September and included two scientists and a professional pilot. Two flight paths (covering the eastern and the western part of the gulf) were followed, each path being completed in about 5 hours. The total area covered is broadly delimited by (3°5'E, 5°2'E) and (42°N, 43°3'N). Each school spotted during a survey day, as well as each daily track, were precisely located through a GPS and a GIS-equipped laptop (for more details on the aerial surveys, see Fromentin 2002). The aerial surveys yielded a total of 80 juveniles BFT schools in 2000 for a number of 6 flight days, 77 in 2001 for a number of 8 flight days and 54 in 2002 for a number of 16 flight days. Field validation was conducted by using purse seiners log-books and involved professional spotters (Fromentin 2002).

Remote sensing data

High-resolution data sets covering the ocean's surface include measurements from various types of sensors (e.g. radiometers, scatterometers, radars). Of the most popular in fisheries oceanography are the Sea Surface Temperature (SST), and the Ocean Colour (OC). SST is

derived from brightness temperature using AVHRR, while OC is calculated from backscattering light level using SeaWiFS. SST and OC have been mostly used in statistical models of the ocean, as tracers of its dynamics and indices of patchiness (Mahadevan & Campbell 2002), or within data assimilation schemes for ecosystems models (Hofmann & Friedrichs 2001). Beyond visual inspection, their main use in fisheries oceanography is to compensate the lack of coincident physical measurement of a given fishing operation (e.g., Bigelow et al. 1999, Maury et al. 2001). Various space/time resolutions are available for these two sensors, from 1.1 km (Local Area Coverage data, or LAC) to 4 km (sub-sampled Global Area Coverage, or GAC) and 9 km. We used here the highest available resolution, i.e. 1.1 km. The SeaWiFS LAC daily raw radiances were downloaded from NASA's Distributed Active Archive Center (URL: <http://eosdata.gsfc.nasa.gov/data/dataset/SEAWIFS/>). Level-2 AVHRR SST maps were supplied by the CMS of Lannion (France). The ground resolution of this product is 2 km (Brisson et al. 2001). A total of 76 SST images (30 In 2000, 29 in 2001 and 17 in 2002) and 49 OC images (19 In 2000, 21 in 2001 and 9 in 2002) were collected to span the whole survey period. Images spoiled by clouds and nebulosity were discarded.

Data merging

All images from both sources were remapped on a Mercator grid with a common resolution of 1/100 of a degree. Daily maps of spotted schools and remote sensing images were aggregated for illustrating purposes in figure 1, 2 and 3. Quantification studies were all conducted on the daily observations, since the merging process may have no ecological meaning.

METHODS

SeaWiFS data L1-L2 processing

A Level-1 to Level-2 process chain was set up using the SEADAS software (Fu et al. 1998). We used the OC-4 algorithm (O'Reilly et al. 1998) to derive type-a chlorophyll concentration from the raw radiances:

$$C(\text{mg.m}^{-3}) = 10^{(a_0+a_1*R+a_2*R^2+a_3*R^3)} + a_4, \text{ with } R = \log\left(\frac{\text{SUP}(Rrs443, Rrs490, Rrs510)}{Rrs555}\right)$$

with Rrs denoting the raw radiance at a given wavelength. OC-4 is less accurate in coastal (“Case II”) waters, than in oceanic (“Case I”) waters, because of high back-scattering of light due to suspended matter. OC-4 tends therefore to overestimate chlorophyll concentration in coastal waters (Gohin et al. 2002). However, this limitation is not so critical for our purpose because: (1) BFT schools are detected offshore (see figure 1) and (2) we were mainly interested in image processing and fronts detection.

Fronts enhancement technique

An automatic front enhancement algorithm can prove to be valuable in the objective assessment of the so-called “tuna-front” association. Various edge detection techniques exist in the literature, involving isotherm contouring, empirical orthogonal functions (Chiswell 1994), surface-fitting (Shaw & Vennell 2000), convolution with kernels (Simpson 1990, with a reference to Canny 1986), and specifically designed algorithms (Cayula & Cornillon 1995). Systematic analysis of AVHRR images resulted in databases of ocean fronts (Ullman & Cornillon 1999), and improved knowledge of frontal dynamics (Miller 2000). Still, frontal analysis on remote sensing images remains rare in fisheries science and is mostly used in applied fisheries oceanography to support fishing activities.

We implemented here the Canny-Derliche filter (Canny 1986) using the Interactive Data Language (IDL). This choice was mainly driven by its simplicity and a wide use in image processing. The method basically involves 4 steps:

- Image smoothing through the convolution with a gaussian filter,
- Gradient computation in both (x and y) directions,
- Non-maxima suppression across the edges,
- Hysteresis thresholding along the edges.

An additional step is invoked to remove fronts smaller than a given length. The parameters involved are the width of the gaussian filter, the width of the gradient estimator, and the low/high hysteresis thresholds. Various values of these parameters were tested empirically. Daily maps of fronts distribution were archived and summed up to produce maps of fronts density over the survey period (figure 3).

Spatial analysis methods

A variety of methods, which come from various research fields, such as statistical mechanics in physics, oceanography, epidemiology, plant ecology or mining statistics, are now available for studying spatial processes in ecology. A detailed review can be found in recent methodological papers (e.g. Fortin et al. 2002). Measures of spatial structure can be roughly separated in two types (Dale et al. 2002): methods dealing with the spatial autocorrelation of gridded or continuous fields of data (e.g., correlograms and variograms) and methods used in the frame of point pattern analysis (distance-based indices where precisely located particles or individuals are considered, e.g., Ripley 1976). Both are applied in this study and discussed.

Estimating local average

Point processes analysis comprises various tools to study the first and second moments of the spatial distribution of individuals (Diggle 1983). Its first-order property is the intensity (or density surface), e.g. the mean number of events per unit area at point s :

$$\lambda(s) = \lim_{ds \rightarrow 0} \left\{ \frac{E(Y(ds))}{ds} \right\}$$

where $Y(ds)$ is the number of individuals per area ds . For a stationary (homogeneous) Poisson process, $\lambda(s)$ will be constant over the study area, so $\lambda(s)=\lambda$. Classic methods for estimating the intensity usually involve quadrat counting. However such a method induces a loss of information at scales smaller than the quadrat size, and interpolative autocorrelation, especially for small data sets (Fortin et al. 2002). We therefore introduce kernel smoothing, a non-parametric method to estimate the density of individuals:

$$\hat{\lambda}_b(s) = \sum_{i=1}^n \frac{1}{\tau^2} k\left(\frac{s-s_i}{b}\right)$$

where k is a user-specified kernel of bandwidth b . Kernel smoothing is an interesting way to study the spatial distribution of free-ranging animals. By characterizing k through a gaussian function, $\lambda(s)$ is given the following ecological meaning: it can be seen as the probability P of finding an individual near any spatial locus after a given period of time t , under the assumption of an unbiased random walk in a homogeneous environment (Gardiner 1983).

$$P(x, y, t) = \int_{-\infty}^{\infty} P(x, y, 0) \cdot K_D(x, y, t) dx dy, \text{ where } K_D(x, y, t) = \frac{1}{\sqrt{2Dt} \cdot \sqrt{2\pi}} \exp\left[-\frac{1}{2} \cdot \frac{x^2 + y^2}{2Dt}\right]$$

K_D is a gaussian kernel of standard deviation $\sqrt{2Dt}$, with D the diffusion coefficient and t the integration time interval. In practice, this integral is computed as a convolution product in the Fourier space. The density estimate will inherit the properties of the kernel, but its choice is not as critical as the choice of the window width (Stoyan & Stoyan 1994). The bandwidth of

the gaussian kernel is related to the diffusion coefficient D , which can be approximated from the average speed of a randomly-walking individual as: $D = \frac{\Delta x^2}{2\Delta t}$, where Δx is the average displacement over time interval Δt . The resulting intensity surface can be interpreted as a potential “search area”. Ultrasonic telemetry applied to juvenile BFT and archival tag on adult BFT provided several records of swimming speeds (Brill et al. 2002, Lutcavage et al. 2000). According to these studies, the average speed is about 6 km/h, yielding a diffusion coefficient of 18 km²/h. The standard deviation of the daily redistribution kernel used in this study is therefore set to $\sqrt{2Dt} \approx 30$ km (with t=24 hours).

Estimating spatial variance

The semivariogram (Matheron, 1971) is used in geostatistics to describe how samples are related with distance. An estimate of the semivariogram is:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} (Z(X_i) - Z(X_j))^2$$

where $N(h)$ is the number of pairs $Z(X_i)$ and $Z(X_j)$ separated by a distance h . Geostatistics are used when several sampling/recording stations cover the area of study, or, as in plant ecology, when this area is subdivided in smaller objects (e.g. blocks or quadrats). The quantity of interest is then the number of individuals per unit block. In the case of animal ecology, mobile individuals (e.g. travelling between quadrats) may bias the spatial statistics. More generally, this new arrangement in the data induces interpolative autocorrelation with a scale similar to the size of each quadrat (Fortin et al. 2002). The same problem can be addressed when considering smoothed density surfaces, and most types of count data.

The lagrangian, or point process, framework appears more suitable for our purpose (Diggle 1983). However, linking spottings data (e.g. point-like data) to the “oceanscape” (e.g. gridded remote-sensing fields in the present study) appears problematic since the two processes are

different in essence. We therefore firstly confronted spotted schools to delimited fronts and eddies, then we analysed location-based data (fish schools) in relation to density-based data (gridded fields). Both can be achieved in the point process framework.

The second-order property is the number of events by pairs of areas:

$$\gamma(s_i, s_j) = \lim_{ds_i, ds_j \rightarrow 0} \left\{ \frac{E(Y(ds_i)Y(ds_j))}{ds_i ds_j} \right\}$$

For a stationary Poisson process $\gamma(s_i, s_j) = \gamma(s_i - s_j) = \gamma(d)$, e.g. γ depends only on the direction and distance between s_i and s_j . An equivalent of this quantity in geostatistics would be the semi-variogram. A second-order statistic in point process is given by Ripley's K -function (Ripley 1976), widely used in ecology: it measures spatial structure at different ranges, under the assumption of homogeneity and isotropy. An estimate of K is provided by:

$$\hat{K}(d) = \frac{1}{\lambda} \sum_{i=1}^n \sum_{j=1}^n (\delta_{ij}(\|x_i - x_j\| \leq d)) / n, \text{ where } \delta_{ij} \text{ is } 1 \text{ if } \|x_i - x_j\| \leq d, 0 \text{ otherwise.}$$

An edge effect correction is necessary near the boundaries of the study area (Ripley 1977). Under Complete Spatial Randomness (CSR), such as Poisson process, the expected $K(d)$ is equal to $\pi \cdot d^2$. A variance-stabilized version of \hat{K} is the Besag function (Besag 1977):

$$\hat{L}(d) = \sqrt{\hat{K}(d) / \pi} - d$$

$L(d)=0$ is equivalent to $K(d)=\pi \cdot d^2$ and indicates Complete Spatial Randomness, $L(d)>0$ indicates more clustering than expected from a Poisson process (e.g. CSR), $L(d)<0$ indicates more repulsion than expected.

Point pattern analysis in ecology has a one-to-one equivalent in statistical physics, namely Smoothed Particles Dynamics, where similar techniques are used (e.g. pair correlation functions and kernel estimates). Another point is that classic geostatistical methods can be used on the modelled intensity surface, (as well as spatial indices on the quadrat counts), but

the structural (interpolative) autocorrelation induced by smoothing may severely bias the second moment estimate.

Accounting for heterogeneity when estimating spatial variance

Spatial heterogeneity is fundamentally inherent to natural ecosystems. The assumption of homogeneity and isotropy is thus violated, and the system may be characterized by spatial autocorrelation driven by the environment (e.g. resource availability), or due to intraspecific interactions (e.g. schooling or swarming). Therefore, spatial aggregation can arise from dependence with a patchy underlying process (for example food - plankton or nekton), without necessarily having (second-order) interaction mechanisms between individuals (the reverse is also possible). It is thus important to allow for heterogeneity when specifying an underlying model. Following Foldager & Pedersen (2002), we modelled the spatial distribution of observed fish schools, conditional on a model taking in account process (e.g. biotic) heterogeneity, environmental variability, and possible interaction effects between fish aggregates. In practice, the intensity $\lambda(x)$ is fitted to the point data as a linear combination of external variables, using the method of maximum pseudolikelihood (Besag 1975). The external variables considered in this study are: the sea surface temperature, the chlorophyll concentration, the distance to nearest thermal or ocean colour front. Compared to the homogeneous Ripley's K which assumes a Poisson random distribution, the heterogeneous version of Ripley's K (Baddeley et al. 2000) assumes a user defined intensity surface $\lambda(x)=f(\text{environment})$ and can be written as:

$$\hat{K}_{heterog}(d) = \sum_{i=1}^n \sum_{j=1}^n \frac{(\delta_{ij}(\|x_i - x_j\| \leq d))}{\lambda(x_i)\lambda(x_j)}$$

$K_{heterog}$ can reveal second-order aggregative properties, conditional on the environment (the Complete Spatial Randomness model is in this case an heterogeneous Poisson process of

intensity $\lambda(x)$). In other words, the question we investigated is: “are the fish schools more or less aggregated than expected when linking them to environmental descriptors?”. Point process statistics appear well suited to our purpose, compared to Generalized Linear Models (which are good at choosing explaining models, but are likely to fail too if not enough data is available), and geostatistical methods (which are generally useful in extracting the spatial structure of objects, but are inefficient in modelling the underlying process of the observed patterns). Still, the goodness-of-fit of point pattern models is highly sensible to the specification of the intensity surface, which requires a good a priori knowledge. All statistics were implemented using the R language (Ihaka & Gentleman 1996), while point processes were analysed using the Spatstat package (Baddeley & Turner 2002).

RESULTS

Spatial structure of BFT schools patterns

Kernel density estimates between 2000 and 2002 show a heterogeneous distribution over the survey area: most schools were irregularly located offshore in the southern area of the survey (Figure 2). Classic geostatistical methods were performed on the daily smoothed density surfaces. Empirical variograms exhibited a range at around 40 km, with roughly linear increases for shorter distances (figure 4). However, the variograms were highly variable from day-to-day and secondarily from year-to-year: both total variance and range varied extensively. While the variance appears clearly related to the number of spotted BFT schools (i.e., the more detections, the greater the variance), changes in range appeared more sensitive to the smoothing kernel bandwidth, as indicated by additional sensitivity analysis. Spatial autocorrelation curves rapidly decreased to low values around 40 km (a range similar to these of the variograms), but null values were not obtained until 80 km (figure 4). By comparing

these curves to the autocorrelation of a gaussian signal of squared-error $D=18$ (dashed lines), we could get information about the additional features present in the data (e.g. remaining after smoothing). The main differences between the two curves appeared at around 30 km, which roughly corresponds to the bandwidth. Note that these curves were estimated by Fourier transform of the data and were therefore smoother than the empirical variograms. Variance-stabilized Ripley's K were positive for all distances, which indicate aggregation/clustering in the stationary case (figure 4). All three annual curves exhibited peaks between 10 and 80 km, indicating more aggregation than expected from a random Poisson process at these ranges.

A sensibility analysis was conducted using various kernel bandwidths (i.e. school swimming speeds of 3, 6 and 9 km/h). As expected, the small scale structure of BFT schools patterns tended to fade when using large bandwidths, whereas low bandwidths lead to high spatial variability (e.g., roughness). An average speed of 6 km/h appeared as a satisfying trade-off for the interpretation of the correlation curves and furthermore sensible from a biological viewpoint.

Sea surface dynamics in the Gulf of Lions

The frontal dynamics in the Gulf of Lions were derived from daily remote-sensing images, collected during the aerial survey periods (i.e., mainly. august and September 2000, 2001 and 2002, see figure 3). The observed dynamics therefore correspond to the summer season. The mean surface temperature of the water masses in the study zone is comprised between 21 and 22 °C. The mean chlorophyll concentration exceeds slightly 0.3 mg/m^3 , with a peak in the middle of area. A strong North-South gradient can be seen in surface temperature, whereas the ocean colour as well as the front density maps displayed more complex and patchy structures. The extension of the Rhône plume is well delimited in the mean chlorophyll field. In general frontal zones are more visible in ocean colour, and is well covered by the survey

path (figure 3). The Gulf of Lions scene is regularly covered by high and low altitude clouds, making conventional radiometers often inefficient. Cloud-free images display characteristic patterns of temperature and ocean colour delimited by fronts, jets and eddies, as seen for example in figure 5. This area extends from the Rhône river mouth (N-E corner) to Cape Béar (S-W corner), following roughly the 1000 m isobath. This is a clear signature of the discharge of water masses with high concentrations of Particulate and Dissolved Organic Matter and nutrients, interacting with the general dynamics of the Liguro-provençal current. This kind of pattern is often seen when the northern wind jet (known as the Mistral and Tramontane) weakens, making the Rhône plume clearly visible due to its cold signature and a high nutrients content.

Superimposing locations of BFT schools on environmental maps

Plotting BFT schools spotted during 2000, 2001 and 2002 on the corresponding ocean colour and temperature maps clearly indicated that the fish distributions are not random. Preliminary visual inspections revealed that they are in general close to surface structures such as the Rhône plume, and at a more local scale to thermal and ocean colour fronts. The scene corresponding to september, the 9th, 2000 is provided here as an illustrative example (figure 5). Few isolated schools were spotted during this specific day: most of them were clustered in two major aggregates located in relatively cold, nutrient-rich waters originating from the Rhône river mouth. They also appeared tightly associated to some straight thermal/ocean colour fronts crossing the flight transects. To assess the significance of this apparent association, frequency histograms of surface layer temperature, chlorophyll-a concentrations were computed using their values of the day at all BFT schools locations. Histograms of distances to nearest thermal and ocean colour fronts were also computed and displayed on figure 6. Then, the empirical distributions of these four variables were gathered (e.g. using the

values of the day at all grid points in the whole survey area). Comparing those two distributions allowed us to test if the BFT schools spotted during a specific day gathered at particular values of oceanographic variables, or if these values at each school location were only reflective of the oceanographic situation in the Gulf of Lions.

The BFT schools apparently occupied very narrow ranges of temperature (between 20.5 and 21.5°C) and phytoplankton concentrations (0.2 to 0.3 mg/m³). They also showed preference for short distances to ocean colour fronts (between 0 and 10 km) and, to a lesser extent, to thermal fronts. Kolmogorov-smirnov two-sample tests (Sokal and Rohlf 1969) confirmed that the distributions of these variables were highly different from their empirical distributions in the whole area. The null hypothesis H_0 was: “BFT schools and external factors are distributed similarly”, while the alternative hypothesis was “the spatial distribution of BFT schools and explaining factors are dissimilar or weakly overlapping”. H_0 was rejected in most cases, but this was less remarkable for the distribution in relation to fronts ($p=2.10^{-8}$ and $p=1.3.10^{-9}$ for sea surface temperature and phytoplankton concentration, respectively; $p=0.05695$ and $p=0.00029$ for distances to nearest thermal front and nearest ocean colour front, respectively). This result can be thus interpreted as follows: sea surface temperature fields, as well as the phytoplankton fields were not sampled randomly by the set of BFT schools locations. They showed preference for narrow ranges of values which appeared to coincide with frontal regions. Those results show that, although BFT schools may have sought precise values of surface temperature and/or phytoplankton, it is more likely that the frontal regions, for this specific day showing strong surface variability, were better described by narrow peaks in SST or OC histograms than the strength of the gradients themselves.

Linking BFT schools distribution to the environment

As an attempt to account for non-stationarity on a daily basis, the inhomogeneous versions of Ripley's K were computed, conditional to a linear combination of external descriptors. September the 9th, 2000 is again shown here for illustration. Four point process models were fitted, always including the distance to flight path as an explaining variable (i.e. to account for patterns induced by the sampling design). The following combinations were tested: (1) a combination of the distances to the nearest ocean colour and thermal fronts, (2) the distance to nearest thermal front along with the surface temperature field, (3) the distance to nearest ocean colour front along with the phytoplankton field, and (4) all the five variables. The intensity surfaces were fitted as third-order polynomials of these variables using the method of maximum pseudolikelihood and the inhomogeneous Ripley's K were then computed, taking each fitted intensity surface as the CSR null model. The variance-stabilized K is expected to be zero at all scales under the CSR hypothesis. Values above zero at a specific scale indicate over-aggregation, whereas values below zero indicate over-spreading.

The goodness of fit of the various models are illustrated by the Ripley's K curves in figure 8. For this specific day, the model including temperature and phytoplankton performed better at short scales (e.g. less than 20 km) than the models depending on the distance to fronts, or the distance to the flight path only. The distance to flight path was an important variable in all models. As expected, the total model (gridded fields + distance to fronts + distance to flight path) performed the better and the K curve displayed values close to zero from scales between 0 and 40 km. As it can be easily seen, the probability surface conditional to the total model is more restricted than those of the other models and the great majority of the BFT schools were distributed in the areas of highest probabilities (figure 7). However, the K curve of the total model displayed slight positive values at around 5 km and became largely negative beyond 40 km, indicating that the model could not explain some spatial structures at very small and large

scales. In other words, the results of the various models indicated that BFT schools patterns were closely related to (or well delimited by) oceanic features, except at the smallest scales (< 10 km) and at the largest ones (> 40 km), where no detection occurred in potentially favourable zones.

DISCUSSION

Hydrodynamics of the Gulf of Lions

The dynamics of the Gulf of Lions has been described in past studies, using remote sensing data and *in-situ* measurements from moored buoys and research cruises, as well as numerical models (e.g. Demarcq 1985, Estournel et al. 1997). This area is under the influence of three dominant processes: (1) strong northern winds inducing coastal upwellings, (2) the massive discharge of nutrient rich waters from the Rhône and (3) offshore branches of the liguro-provencal current. The first assessment of fronts occurrence in the Gulf of Lions was made by Demarcq (1985), using AVHRR derived temperature maps and CZCS derived ocean colour maps. Using new and more powerful algorithms, we here showed that ocean colour fronts are mostly located in the Rhône plume's extension, while thermal fronts appeared more scattered and related to other mechanisms, such as stirring and mixing with the liguro-provencal current. Furthermore, our results showed that thermal and ocean colour fronts were decoupled in space and time, especially offshore. This is in agreement with previous findings on the patchiness of these variables. In fluid dynamics, various energy inputs from thermal and gravitational origins result in convection and advection processes, leading to the formation of spatial discontinuities, where enhanced horizontal and vertical stirring occur. Such surface heterogeneity can be seen through sea surface temperature maps. Biological processes (e.g., reproduction, recruitment, growth) are by nature aggregative, leading to added patchiness in geophysical variables such as ocean colour. It has been shown by Mahadevan & Campbell

(2002) that the characteristic response time of phytoplankton to stirring and nutrients is different to the one of temperature in response to heat fluxes. Phytoplankton may therefore be a patchy variable, with more underlying processes than temperature, which appears to be a more pertinent tracer of ocean circulation. In the special case of remote-sensing, Mueller & LaViolette (1981) stated that ocean colour fronts are suggestive of turbulent diffusion, whereas thermal fronts reveal shear instabilities. Moreover, the AVHRR infrared sensor only deals with temperature at the very surface, whereas pigment concentration is integrated over the upper layer of the ocean. Thus, the position of thermal and OC fronts match only up to a certain extent, both in space and time. The implications of these facts are further discussed.

Interpreting the spatial distribution of bluefin tuna schools

The empirical variograms tend to indicate a spatial structure of the BFT schools at around 40 km, but they varied widely from a day to another. The underlying processes (be them physical or biological) are thus likely to change over time. The second order characteristic of a point process (Ripley's K), as well as the spatial autocorrelation plots, revealed that the fish schools were clustered over a wider range of scales (from 10 to 80 km), indicating more aggregation than expected from a random process. As the biophysical properties of the surrounding water masses may have an influence on the BFT schools, their spatial structure is likely to reflect this relationship.

Superimposing the schools detected by aerial surveys on environmental maps effectively showed that their spatial distribution was linked to some specific oceanic features. This visual inspection was in agreement with the histograms and Kolmogorov-smirnov tests, which showed that the schools were not randomly distributed in relation to the sea surface variables on a daily basis. It was also shown that, while clusters of schools were found mostly in the vicinity of frontal lines, these "preferred" zones were better characterized in general by

narrow ranges of SST and/or phytoplankton concentration. However, we believe that this cannot be interpreted straightly as preferendii: pooling the whole set of data over 2000, 2001, and 2002 yielded different results, as shown in figure 9. Temperature exhibited three peaks at 16, 21 and 24°C (corresponding to a seasonal signal sampled over the different survey periods), whereas chlorophyll concentration exhibited a single peak at 0.25 mg/l. It was not possible anymore to establish any preferendum: the distributions of these values at each school location were not significantly different from their empirical distribution in the entire survey area. This was confirmed by the acceptance of H_0 in the Kolmogorov-smirnov tests ($p=0.102$ and $p=0.303$ for SST and OC respectively). On the other hand, the distances to nearest surface front were still sampled with a preference for small values (roughly less than 10 km): H_0 was rejected for ocean colour fronts ($p=0.028$) and, to a lesser extent, for thermal fronts ($p=0.071$). This is interpreted as follows: when pooling the daily data, local relationships may vanish, while association with frontal features remains significant. Our conclusion is that fronts are ubiquitous features occurring at the sea surface and may be better suited in describing relationships over entire seasons, while optima in SST or OC appear more difficult to characterize since they strongly depend on local conditions (both in space and time). Concerning front detection and characterization, alternative techniques (e.g. dynamics oriented, gradient independent algorithms) may be worth investigating in fisheries oceanography, such as the method developed by Cayula & Cornillon (1995).

BFT schools distributions were in general well determined by the oceanic features (gridded fields and distance to thermal and ocean colour fronts) as soon as the flight path was further taken into account. We were roughly able to model correctly the spatial distribution of BFT schools and to reproduce what could be seen from a simple visual superimposition of the sightings on environmental maps. The spatial distribution of the BFT schools appears thus clearly non-stationary since: (1) sampling autocorrelation is induced by the survey design (i.e.

the transect sampling, although the number of sightings decreases exponentially in relation to the perpendicular distance of the flight) producing a non-fully mapped data set and (2) BFT schools distributions appeared, in general, to be linked to the ocean “landscape”, which displays a variety of scales in its surface variability (e.g. long and small range gradients). This latter process in itself can induce autocorrelation and non-stationary distributions in marine species, particularly if the distribution of the spotted schools does not span the whole range of environmental variables (i.e., the ergodic hypothesis is not valid here).

However, several results also clearly pointed out that some spatial structures still were unexplained by oceanic features, especially at very small scales (< 10km) and at larger scales (> 40km). Such an outcome could either be explained by: (1) other environmental factors than these used in this study are of importance (e.g., the spatial distribution of tuna forage), (2) the Ripley’s K could be biased at larger scales because the number of comparisons decreases as the lag increases (this point would not, however, explain over-aggregation at small scales), (3) fast moving fronts may be difficult to spot, as the time of the survey flight and the recording time of the images can differ of several hours, and (4) some biological and/or behavioural processes occurred and led to a departure from aggregation modelled by environmental constraints. For instance, the migration patterns of juveniles BFT within the study area (i.e. the Gulf of Lions), which remain largely unknown, could explain such a departure at large scales. At smaller scales, over-aggregation could also be due to interactions between BFT juvenile schools, leading to larger than expected gatherings in specific zones. Another major cause for clustering on feeding grounds may be the spatial distribution of their prey, such as clupeids, that are strongly affected by oceanographic variability. Indeed, the interplay between density-independent responses to environmental features and density-dependent social behaviour is likely to produce complex patterns over a variety of scales (Flierl et al. 1999, Steele 1989).

Potential benefits of an association between tunas and ocean fronts

Our results showed that surface fronts are important in explaining partially the spatial distribution of juvenile bluefin tuna schools. Ocean fronts are synoptic structures (or “features”) that form at the interface of neighbouring water masses with different oceanographic characteristics. They are associated with convergent flow, which results in the trapping of material in the frontal region. The advection of nutrients in conjunction with enhanced turbulence form a suitable basis for primary production (e.g. Longhurst 1998) and leads to enrichment, concentration and retention of phytoplankton and zooplankton at fronts (e.g. Bakun 1996). Many marine fishes therefore concentrate in fronts during their early life stages (Lough and Manning 2001). Association of top predators with coastal or permanent geostrophic fronts has also been documented in Laurs et al. (1984), Fiedler (1987), and Santos (1994). Inagake (2001) and Lutcavage (2000) describe qualitatively such an association in the case of the Pacific and the Western Atlantic bluefin, respectively. Several hypothesis were proposed to explain such an association, among them the use of fronts as preferred foraging areas, the benefits of sharper temperature gradients for thermal regulation, or a loose association with trapped floating material as meeting points or navigation cues. In Kirby et al. (2000), explicit modelling of tuna behaviour lead to the conclusion that swimming and diving patterns of individuals at fronts are influenced by a diversity of external factors (ambient temperature, prey availability...) driving internal factors (thermal stress, stomach fullness...). Still, accurate field data is needed to finely tune and validate such explanatory model. Our results indicate that the association between juvenile BFT and frontal meanders may be linked to feeding behaviour since tuna schools did not show striking preference for the warm or cold side of thermal fronts, while enrichment and mixing of water masses are key processes in the Gulf of Lions. Foraging in the vicinity of a phytoplankton front may appear

as a trade-off between prey availability and hunting efficiency owing to water clarity. This is in agreement with previous findings stating that the relationship between tuna aggregates and frontal meanders is most probably indirect and trophic-related, which has been confirmed by observation through aerial spotting in this study (Fromentin, pers. comm.). Advected material at fronts can provide favourable feeding grounds to small clupeids, which are in turn sought by bluefin tuna. Direct processes such as thermal regulation may be less relevant (Brill et al., 1999 and Boustany, 2001). The identification and understanding of these indirect processes is a challenging task for explaining small scale behaviour of tuna.

CONCLUSION AND PERSPECTIVES

In this study, we were able to provide some statistical evidence of a close association between schools of juvenile bluefin tuna and oceanic features, which has been documented for this species in other areas but never quantified before. Point pattern analysis and modelling offered a convenient way of linking the spatial distribution of discrete events (BFT schools) and sea surface variability. This is of importance for future work, since the role of these oceanic features must be explicitly modelled, or at least properly parameterised when describing the distribution of pelagic species. However, the generalization of these results is limited by problems inherent to measurements and survey design (e.g. sensitiveness of radiometers and spotters to cloud covering), the lack of objective definition of frontal regions, and the difficulty of inferring dynamical processes from superimposed static patterns or “snapshots”. Null models of heterogeneous spatial aggregation become then hazardous to specify. We believe that, in such a lagrangian framework, dispersion statistics can benefit from individual dynamics: further work will be conducted on fish behaviour, as revealed by pop-up archival tagging in the Mediterranean Sea. The ultimate goal of this approach is to provide a basis for the stochastic parameterisation of fish movement in environmentally explicit population models.

ACKNOWLEDGEMENTS

Our thanks first go to H. Farrugio (IFREMER Sète) for his work on the aerial surveys and to G. Larnicol and G. Weller (CLS) for their help regarding spatial oceanography issues. The authors would like to thank the SeaWiFS Project (Code 970.2) and the Goddard Earth Sciences Data and Information Services Center/Distributed Active Archive Center (Code 902, Goddard Space Flight Center), for providing SeaWiFS data, as well as the Dundee University HRPT receiver for LAC data. We thank too H. Roquet and P. Le Borgne (Centre de Météorologie Spatiale of Lannion) for providing daily SST data and valuable help in their processing. This work could not have been conducted without the doctoral fellowship fund provided by CLS and IFREMER for FR.

LITERATURE CITED

- Baddeley A, Moller J, Waagepetersen RP (2000) Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54:329-350
- Baddeley A, Turner R (2002) *Spatstat: Analysis of spatial point patterns*. URL: www.maths.uwa.edu.au/~adrian/spatstat.html
- Bakun A (1996) *Patterns in the Ocean: Ocean Processes and Marine Population Dynamics*. University of California Sea Grant, San Diego, California, USA, in cooperation with Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico
- Besag J (1975) Statistical analysis of non-lattice data. *The Statistician* 24:179-195
- Besag J (1977) Efficiency of pseudo-likelihood estimation for simple Gaussian fields. *Biometrika* 64:616-618
- Bigelow KA, Boggs CH, He X (1999) Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. *Fish. Oceanogr.* 8:178-198
- Boustany A, Marcinek DJ, Keen J, Dewar H, Block BA (2001) Movements and temperature preferences of Atlantic bluefin tuna (*Thunnus thynnus*) off North Carolina: A comparison of acoustic, archival and pop-up satellite tags. In: Sibert JR, Nielsen JL (eds) *Electronic tagging and tracking in marine fisheries*. Kluwer Academic Publishers, Dordrecht, The Netherlands, p 89-108
- Brill R, Lutcavage M, Metzger G, Bushnell P, Arendt M, Lucy J, Watson C, Foley DG (2002) Horizontal and vertical movements of juvenile bluefin tuna (*Thunnus thynnus*), in relation to oceanographic conditions of the western North Atlantic, determined with ultrasonic telemetry. *Fish. Bull.* 100:155-167
- Brill RA, Block BA, Boggs CH, Bigelow KA, Freund EV, J. MD (1999) Horizontal movements and depth distribution of large adult yellowfin tuna (*Thunnus albacares*)

- near the Hawaiian Islands, recorded using ultrasonic telemetry: implications for the physiological ecology of pelagic fishes. *Marine Biology* 133:395-408
- Brisson A, Le Borgne P, Marsouin A (2001) North Atlantic Regional Sea Surface Temperature Product Manual Version 1.1, Météo-France, Lannion
- Canny J (1986) A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8
- Cayula JF, Cornillon P (1995) Multi-image edge detection for SST images. *J. Atmos. Oceanic Technol.* 12:821-829
- Chiswell S (1994) Variability in sea surface temperature around New Zealand from AVHRR images. *N. Z. J. Mar. Freshwat. Res.* 28:179-192
- Cowling A, O'Reilly JE (1999) Background to review of aerial survey: 3rd draft 1999 Aerial Survey Workshop, Caloundra, p 61
- Dagorn L, Menczer F, Bach. P, Olson RJ (2000) Co-evolution of movement behaviours by tropical pelagic predatory fishes in response to prey environment: a simulation model. *Ecol. Modell.* 134:325-341
- Dale MRT, Dixon P, Fortin M-J, Legendre P, Myers DE, Rosenberg MS (2002) Conceptual and mathematical relationships among methods for spatial analysis. *Ecography* 25:558-577
- Demarcq H (1985) Applications de la télédétection infra-rouge et visible en océanographie. PhD thesis, Univ. Aix-Marseille II
- Diggle PJ (1983) Statistical analysis of spatial point patterns, Academic Press, London
- Estournel C, Kondrachoff V, Marsaleix P, Vehil R (1997) The plume of the Rhone: numerical simulation and remote sensing. *Continental Shelf Res.* 17:899-924

- Farrugio H (1981) Exploitation et dynamique des populations de thon rouge, *Thunnus thynnus* (Linné 1758) Atlanto-Méditerranéennes. PhD thesis, Université des Sciences et Techniques du Languedoc
- Flierl G, Grünbaum D, Levin S, Olson DB (1999) From Individuals to Aggregations: the Interplay between Behavior and Physics. *J. Theor. Biol.* 196:397-454
- Fiedler PC, Bernard HJ (1987) Tuna aggregation and feeding near fronts observed in satellite imagery. *Continental Shelf Res* 7:871-881
- Foldager L, Pedersen J (2002) Spatial Distribution of Detected Single Fish: An Application of the Log Gaussian Cox Process. Report No. 03, Biometry Research Unit, Department of Agricultural Systems, Danish Institute of Agricultural Sciences
- Fortin M-J, Dale MRT, Ver Hoef J (2002) Spatial analysis in ecology. In: El-Shaarawi AH, Piegorisch WW (eds) *Encyclopedia of environmetrics*, Vol 4. Wiley, p 2051-2058
- Fromentin J-M (2002) Final Report of STROMBOLI. Report No. EU-DG XIV project 99/022, European Community - DG XIV, Brussels
- Fromentin J-M (2003) The East Atlantic and Mediterranean Bluefin tuna management: uncertainties and alternatives. *Scientia Marina* 67:51-62
- Fromentin J-M, Fonteneau A (2001) Fishing Effects And Life History Traits: A Case-Study Comparing Tropical Versus Temperate Tunas. *Fish. Res.*, in press
- Fu G, Baith KS, McClain CR (1998) SeaDAS: the SeaWiFS Data Analysis System The 4th Pacific Ocean Remote Sensing Conference, Qingdao, China, p 73-79
- Gardiner CW (1983) *Handbook of Stochastic Processes for Physics, Chemistry and the Natural Sciences*, Springer-Verlag, Berlin
- Gohin F, Druon JN, Lampert L (2002) A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters. *Int. J. Remote Sensing* 23:1639-1661

- Hofmann EE, Friedrichs MAM (2001) Models: biogeochemical data assimilation, Academic Press, London
- ICCAT (2003) REPORT OF THE 2002 ATLANTIC BLUEFIN TUNA STOCK ASSESSMENT SESSION. Col. Vol. Sci. Pap. ICCAT, in press
- Ihaka R, Gentleman R (1996) R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5:299-314
- Inagake D, Yamada H, Segawa K, Okazaki M, Nitta A, Itoh T (2001) Migration of young Bluefin tuna, *Thunnus orientalis*, through archival tagging experiments and its relation with oceanographic conditions in the Western North Pacific. Bulletin of the National Research Institute of Far Seas Fisheries 38:53-81
- Kirby DS, Fiksen Ø, Hart PJB (2000) A dynamic optimization model for the behaviour of tunas at ocean fronts. Fish. Oceanogr. 9:328-342
- Laurs RM, Fiedler PC, Montgomery DR (1984) Albacore tuna catch distribution relative to environment features observed from satellites. Deep Sea Research 31:1085-1099
- Longhurst AR (1998) Ecological geography of the sea, Academic press, San Diego. 398 pp.
- Lough RG, Manning JP (2001) Lough, R.G., and J.P. Manning (2001) Tidal-front entrainment and retention of fish larvae on the southern flank of Georges Bank. Deep Sea Research II:631-644
- Lutcavage M, Brill R, Skomal GB, Chase BC, Goldstein J, Tutein J (2000) Tracking adult North Atlantic bluefin tuna (*Thunnus Thynnus*) in the northwestern Atlantic using ultrasonic telemetry. Mar Biol 137:347-358
- Lutcavage M, Kraus S (1995) The feasibility of direct photographic assessment of giant bluefin tuna, *Thunnus thynnus*, in New England waters. Fish Bull 93:495-503
- Mahadevan A, Campbell JW (2002) Biogeochemical patchiness at the sea surface. Geophysical Res. Letters 29:32-36

- Mather FJ, Mason JM, Jones AC (1995) Historical document: life history and fisheries of Atlantic bluefin tuna. Report No. 370, NMFS-SEFSC
- Maury O, Gascuel D, Marsac F, Fonteneau A, De Rosa A-L (2001) Hierarchical interpretation of non-linear relationships linking yellowfin tuna (*Thunnus albacares*) distribution to the environment in the Atlantic Ocean. *Can. J. Fish. Aquat. Sci.* 58:1-13
- Miller P (2000) Multispectral front maps for auto-detection of ocean colour features from SeaWiFS and MODIS Oceans from Space, Venice
- Mueller JL, LaViolette PE (1981) Color and temperature signatures of ocean fronts observed with the Nimbus-7 CZCS. In: Gower JFR (ed) *Oceanography from Space*, p 295-302
- Olson DB, Hitchcock GL, Mariano AJ, Ashjian CJ, Peng G, Nero RW, Podestá GP (1994) Life on the edge: marine life and fronts. *Oceanogr.* 7:52-60
- O'Reilly JE, Maritorena S, Mitchell BG, Siegel DA, Carder KL, Garver SA, Kahru M, McClain C (1998) Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research* 103:24,937-924,953
- Ravier C, Fromentin J-M (2001) Long-term fluctuations in the Eastern Atlantic and Mediterranean bluefin tuna population. *ICES J. Mar. Sci.* 58:1299-1317
- Ripley BD (1976) The second-order analysis of stationary point processes. *J. Appl. Proba.* 13:255-266
- Ripley BD (1977) Modelling spatial patterns. *J. R. Stat. Soc.* 39:172-212
- Santos AM (1994) Satellite Oceanography Applications to the Swordfish Fishery off Portugal's Continental West Coast. MS thesis, University of Lisbon, Portugal
- Shaw AGP, Vennell R (2000) A front-following algorithm for AVHRR SST imagery. *Remote Sens. Environ.* 72:317-327
- Sibert JR (2001) Integrating statistical models of tuna movement in relation to fish attractors Deep Rigs and FADs, AFS Southern Division Midyear Meeting, Jacksonville, FL

Sokal RR, Rohlf FJ (1969) Biometry, Freeman, San Francisco, California

Simpson JJ (1990) On the accurate detection and enhancement of oceanic features observed in satellite data. *Remote Sens. Environ.* 33:17-33

Steele JH (1989) The ocean "landscape". *Landscape Ecology* 3:185-192

Stöcker S (1998) Models for tuna school formation. *Mathematical Biosciences* 156:167-190

Stoyan D, Stoyan H (1994) Fractals, random shapes and point fields, Wiley, Chichester

Ullman DS, Cornillon P (1999) Satellite-derived sea surface temperature fronts on the continental shelf off the northeast U.S. coast. *Journal of Geophysical Research* 104:23459-23478

FIGURE CAPTIONS

Figure 1. Juvenile BFT schools sightings in 2000 (a), 2001 (b), and 2002 (c). Bathymetry is shown in black lines. Dots symbolize fish schools locations.

Figure 2. 2000-2002 bluefin schools density in the Gulf of Lions, estimated with a gaussian kernel filter of width 30 km.

Figure 3. a) AVHRR Sea Surface Temperature, b) SeaWiFS chlorophyll concentration, c) thermal fronts density, d) ocean colour fronts density, averaged over the survey seasons (2000-2001-2002).

Figure 4. Spatial statistics for Bluefin schools density in 2000 (left), 2001 (center), and 2002 (right). Top panel shows daily empirical variograms (grey line) superimposed with yearly-average (black line), middle panel shows the spatial autocorrelation of the intensity surface (dark line) along with the spatial autocorrelation induced by the kernel smoothing (dotted line), bottom panel shows the corresponding Ripley's K functions.

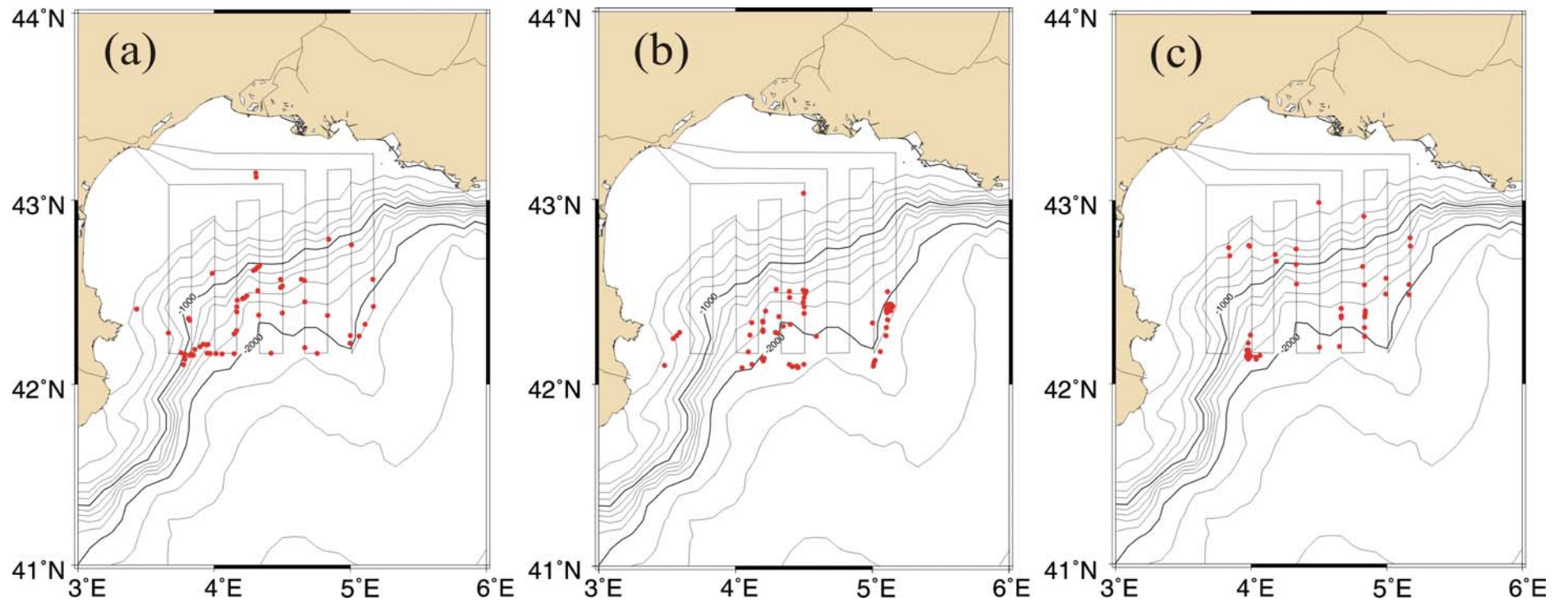
Figure 5. Remote sensing scenes of julian day 18514 (September the 9th, 2000). Crosses mark fish school locations. a) AVHRR Sea Surface temperature, b) thermal fronts, c) SeaWiFS chlorophyll-a concentration, d) ocean colour fronts.

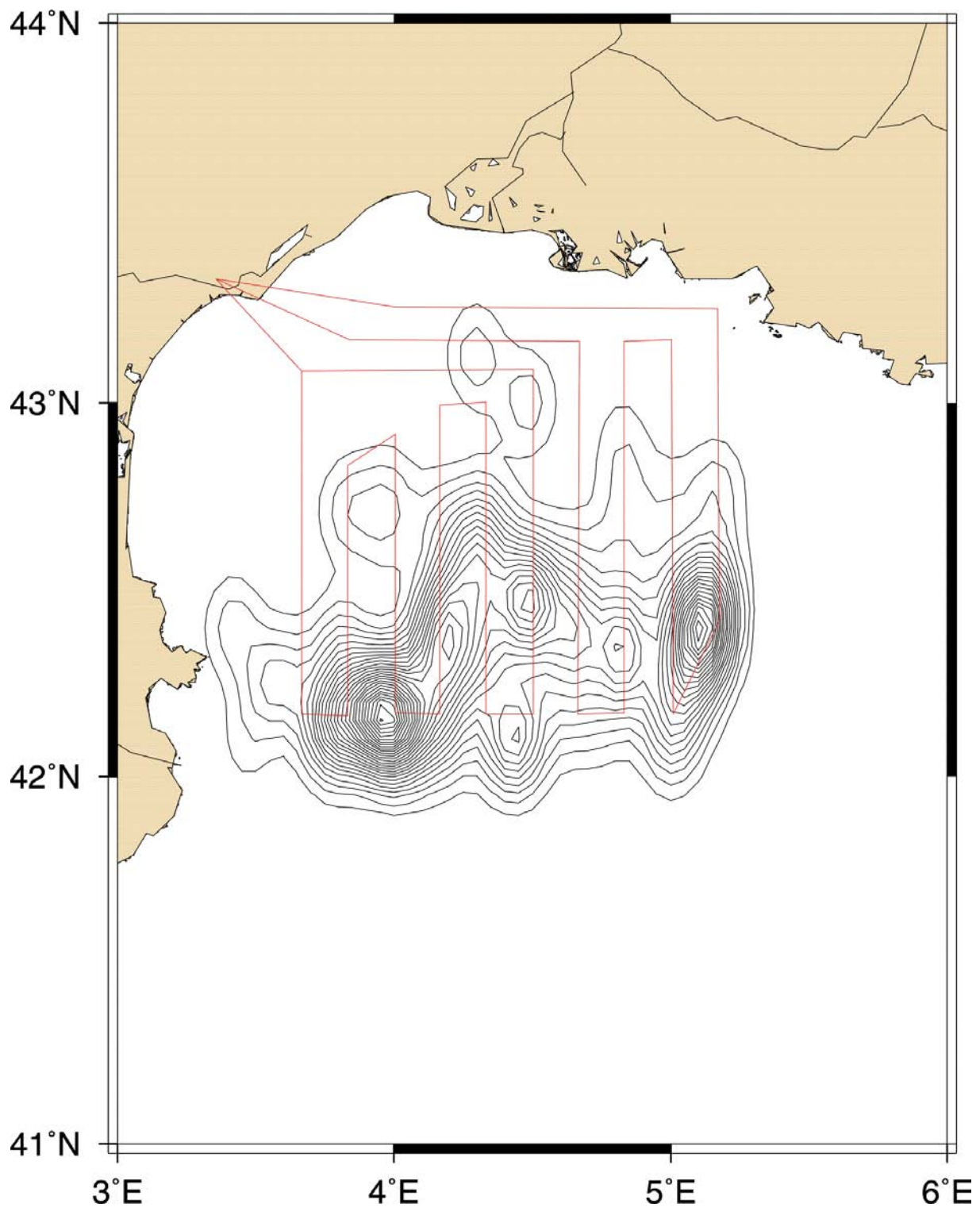
Figure 6. Prior (continuous line) and sampled at each school (vertical bars) distributions of four environmental descriptors, for the survey conducted on september 9th, 2000: sea surface temperature (a), chlorophyll concentration (b), distance to nearest thermal front (c), distance to nearest ocean colour front (d).

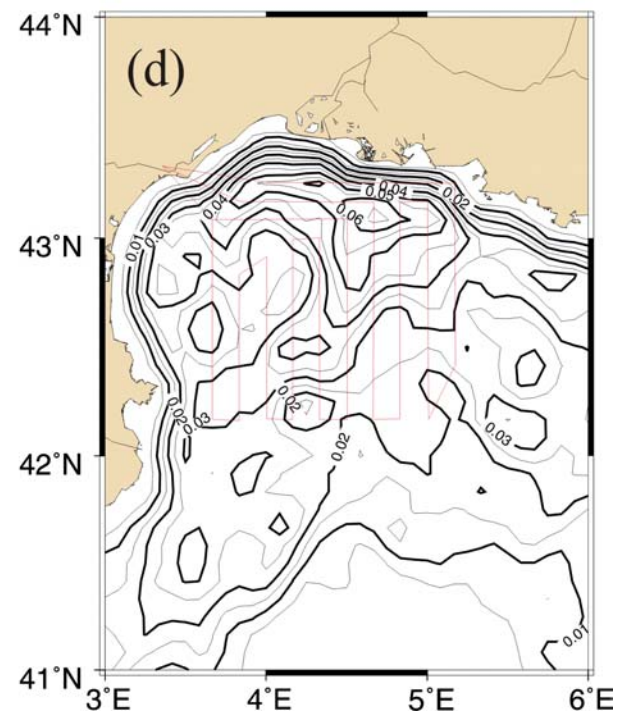
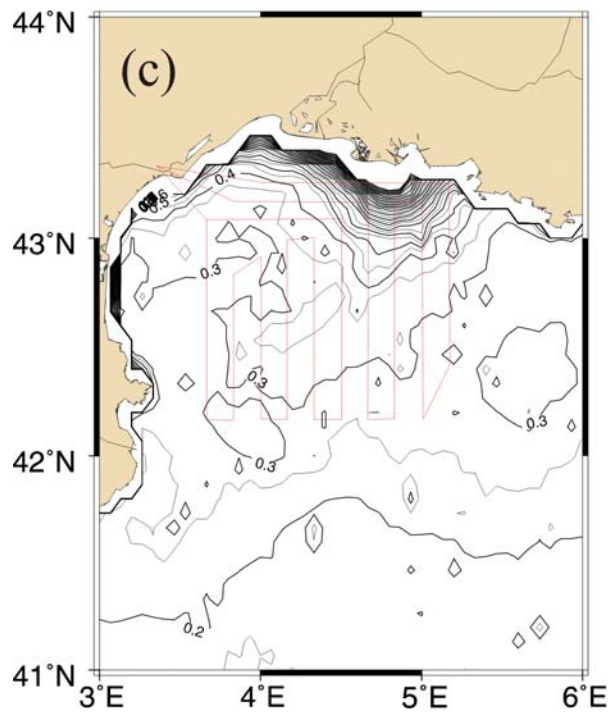
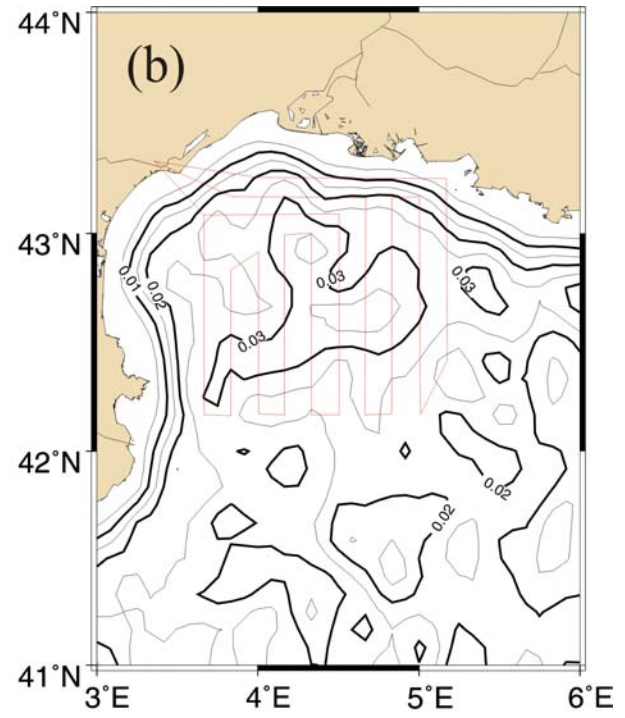
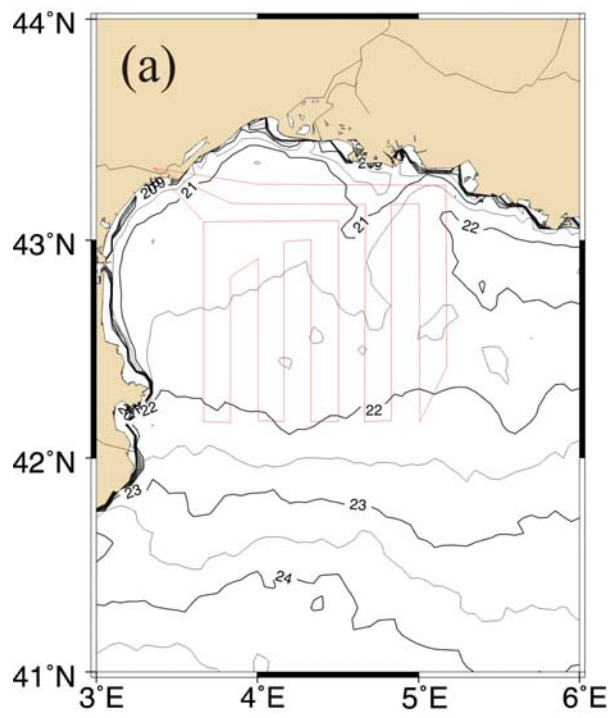
Figure 7. Intensity surface conditional on the distance to flight path (a), temperature and ocean colour field (b), distance to nearest thermal and ocean colour front (c) and all five variables (d). Crosses indicate fish schools spotted on September the 9th 2000. Distance to flight path was calculated as the perpendicular distance between fish school GPS location and on-board record of plane track.

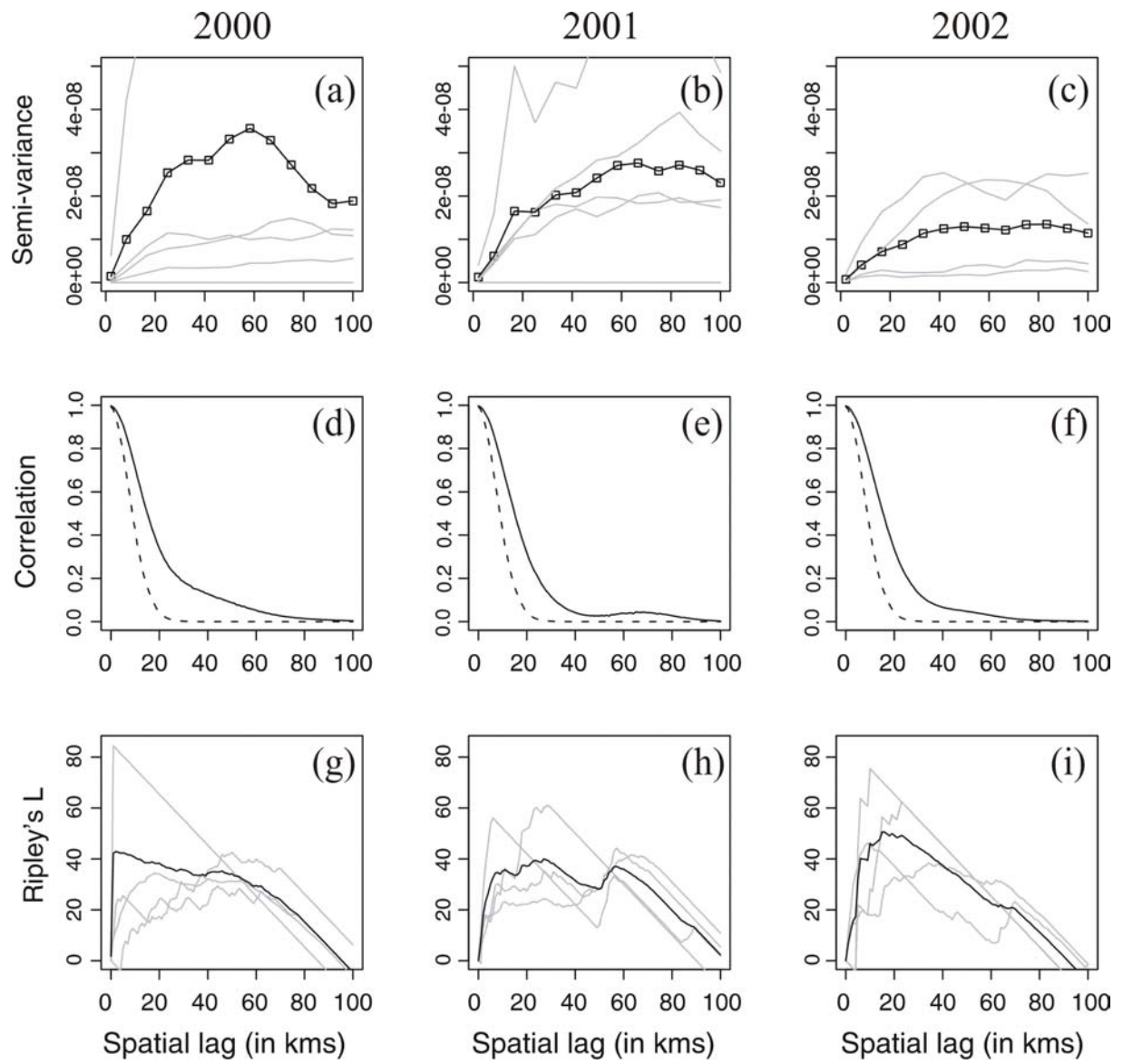
Figure 8. Inhomogeneous Ripley's K conditional on the distance to flight path (a), temperature and ocean colour field (b), distance to nearest thermal and ocean colour front (c) and all five variables (d). Values above zero at a specific scale indicate that the data are clustered in relation to the chosen intensity surface, whereas values below zero indicate over-spreading.

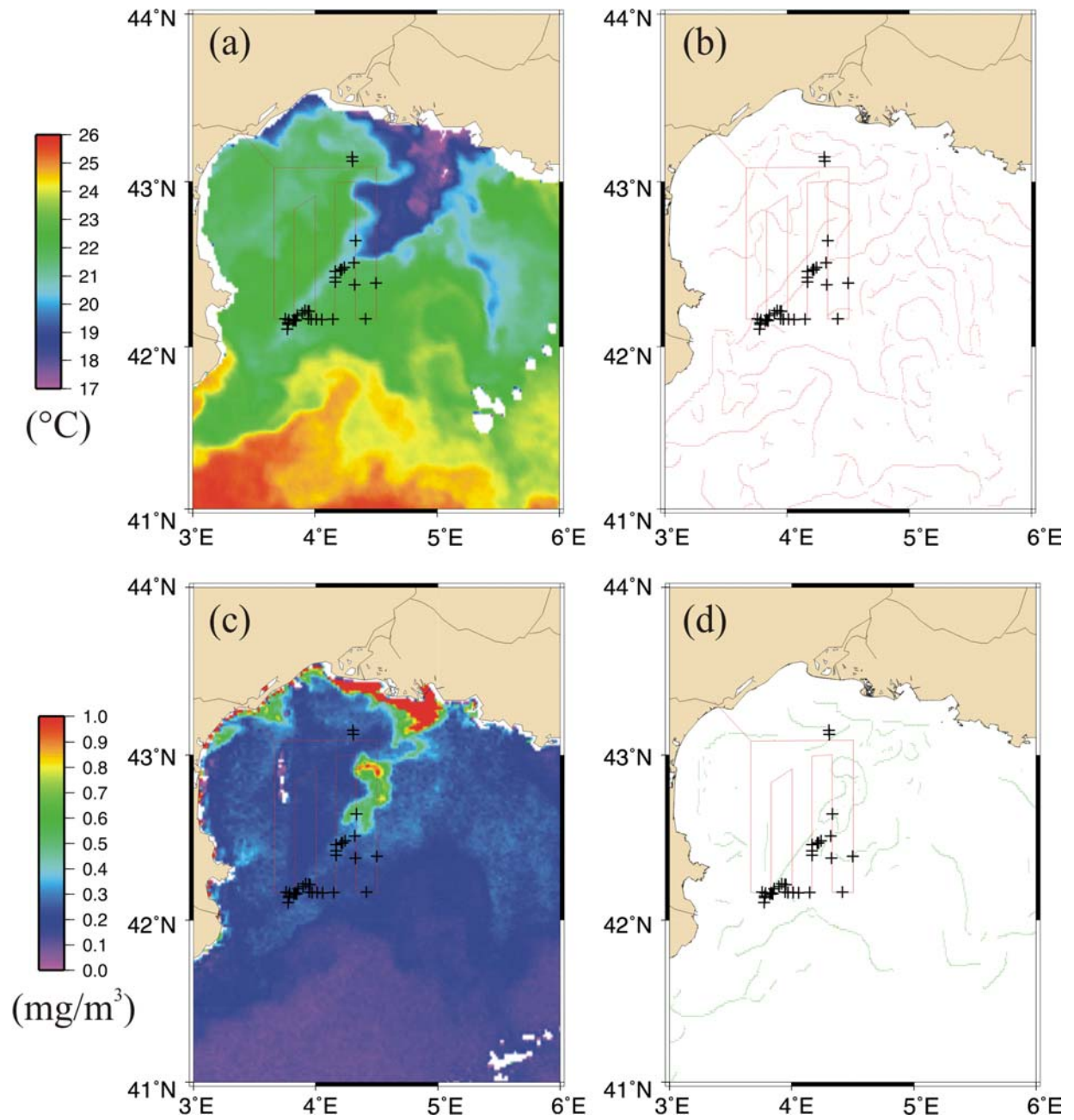
Figure 9. Prior (continuous line) and sampled at each school (vertical bars) distributions of four environmental descriptors, considering the whole survey in 2000, 2001 and 2002: sea surface temperature (a), chlorophyll concentration (b), distance to nearest thermal front (c), distance to nearest ocean colour front (d).

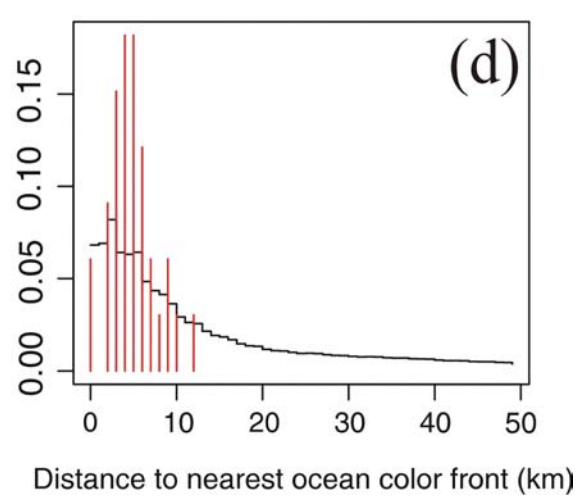
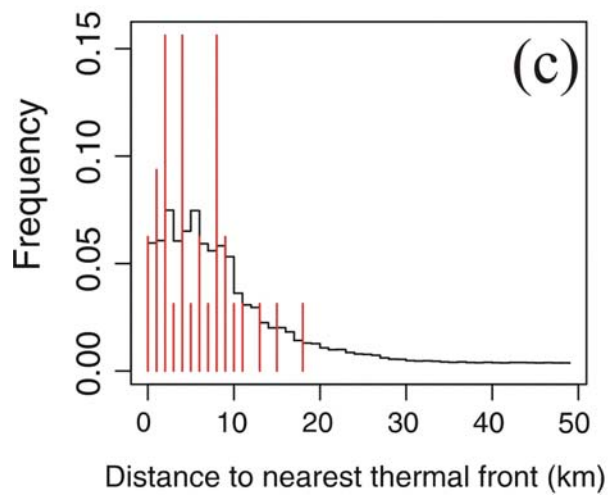
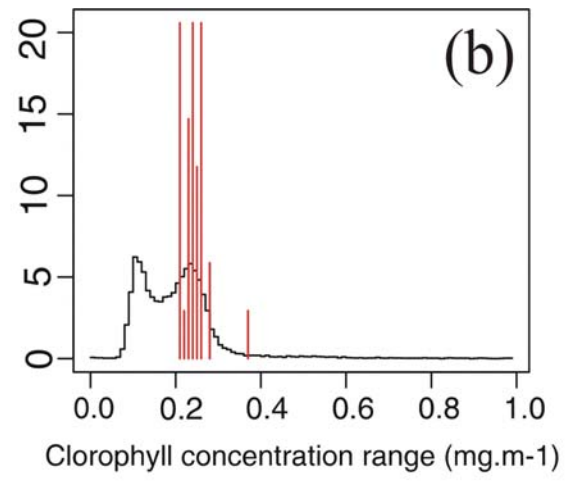
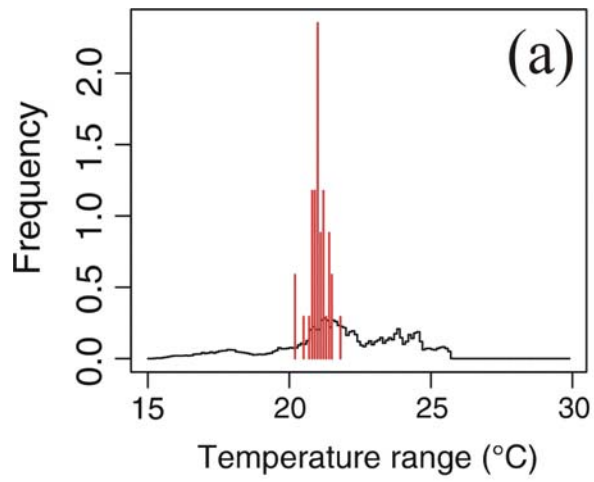


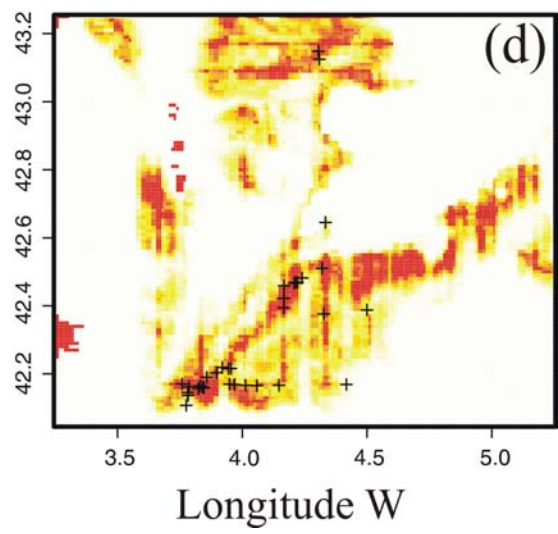
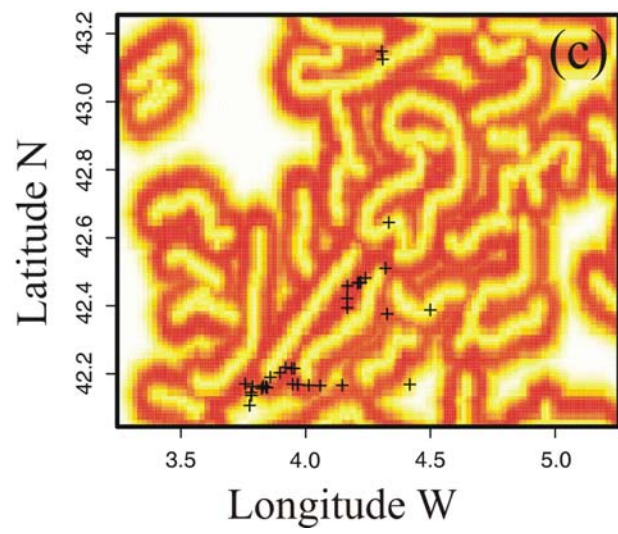
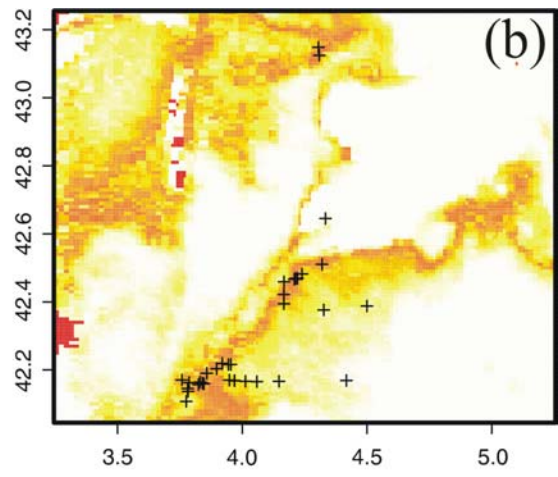
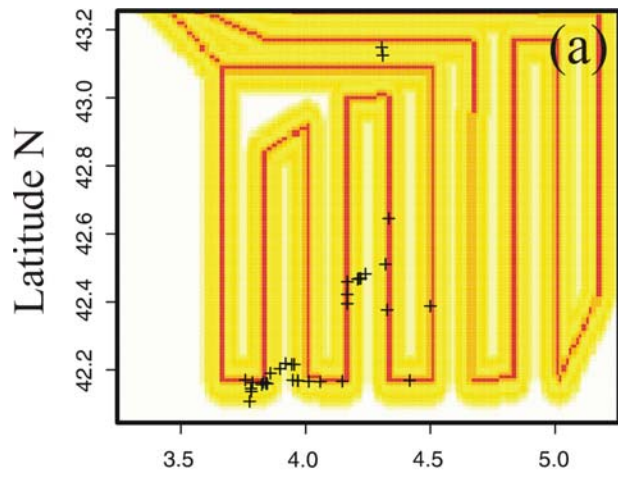


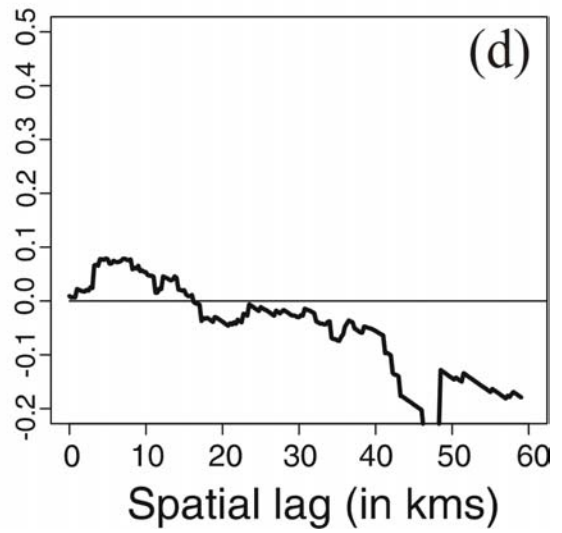
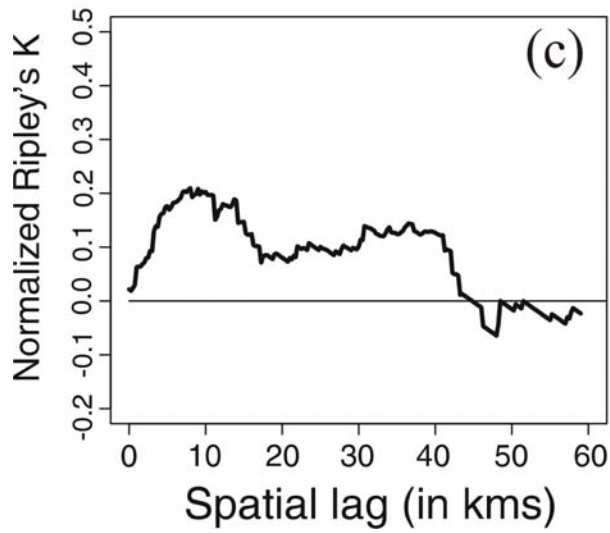
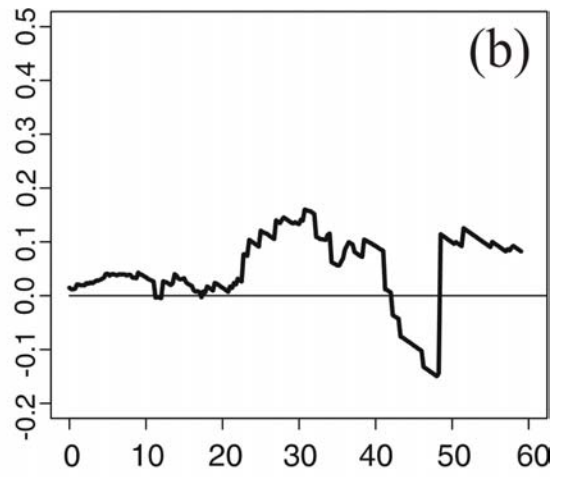
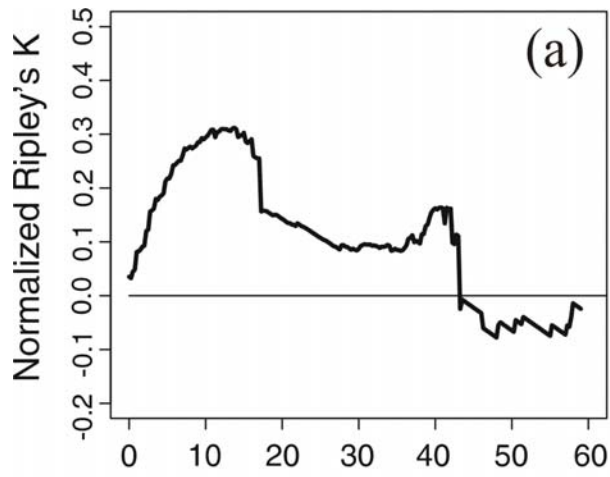


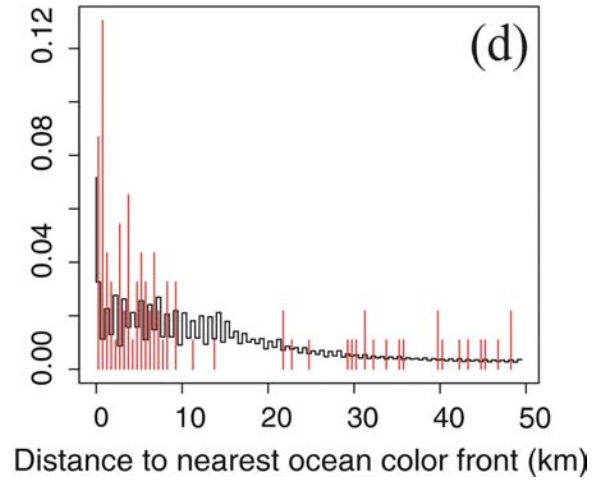
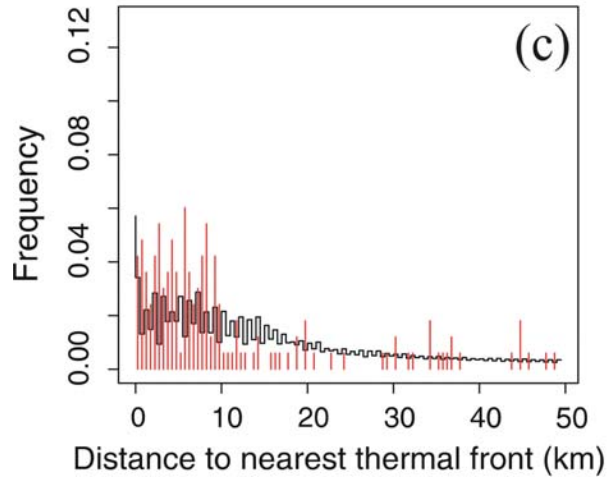
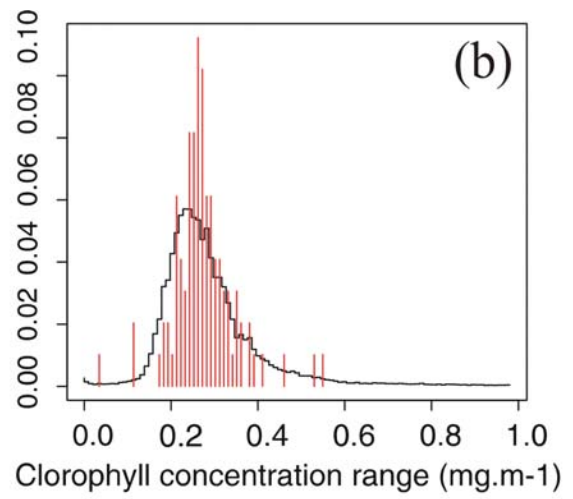
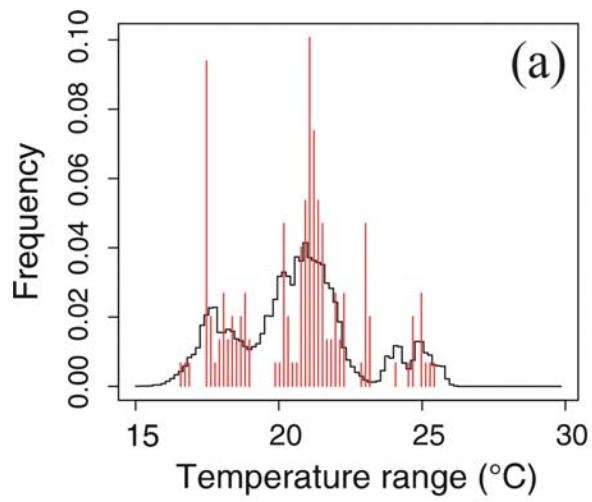












A state-space model to derive bluefin tuna movement and habitat from archival tags

F. Royer^{1,2}, J.-M. Fromentin¹, P. Gaspar²

¹IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean Monnet, BP 171, 34203 Sète cedex, France

²CLS, Division Océanographie Spatiale, 8-10 rue Hermès, 31526 Ramonville St Agne, France

1 **ABSTRACT**

2 Archival tagging provides a unique way to study the spatial dynamics and habitat of pelagic
3 fish. This technique generates lagrangian data of a particular type in marine ecology: although
4 highly informative about processes at different scales (e.g. horizontal movements versus
5 diving behaviour), such data are impaired by location errors and the lack of combination with
6 actual environmental variability. The present paper introduces a framework for modelling
7 bluefin tuna movement in relation to its habitat, using records of light, depth and temperature
8 from archival tags. Based on data assimilation concepts and methods, we show how an
9 explicit formulation of the observation process and the statistics of external variables (e.g.
10 ambient temperature) can improve precision in geolocation. The proposed method is tested on
11 synthetic data: significant reduction (40 to 50%) in the initial Root-Mean Square Error is
12 achieved under different noise scenarios. Assimilating sea surface temperature also allows to
13 perform on-line estimation of a range of observation biases. The performance of the model
14 greatly benefits from the adequate formalisation of different variability sources, and allows
15 potentially to reveal interactions between the fish and its habitat. Using this probabilistic
16 approach, we, however, show that some patterns of interest (e.g. foraging in surface fronts)
17 can hardly be retrieved in a context of large observational and environmental noise.

18

1 INTRODUCTION

2 Northern bluefin tuna (BFT) is a pelagic migratory species distributed in the Atlantic
3 ocean and the Mediterranean sea (Mather, et al. 1995). Its population has undergone a
4 sustained exploitation since the Antiquity, through coastal traps, and recently longlining and
5 purse seining (Fromentin 2003). Like other top predators, its habitat and spatial dynamics are
6 still poorly known, due to the scarcity of direct observations. How large migratory animals
7 make use of the spatio-temporal variability of their physical environment and prey resources
8 is still an open question (Bakun 1996). Knowledge on migration behaviour and long range
9 displacements is not only of particular importance for ecological purposes (see e.g. Ravier
10 and Fromentin 2004), but also for management issues (ICCAT 2003). However, the study of
11 such lagrangian processes remains difficult, both in term of observation and understanding
12 (Nathan, et al. 2003). Advances in this field depend on our ability to link fish behaviour and
13 oceanographic variability over a wide range of temporal and spatial scales. Conventional
14 fishing data (e.g. catches or Catch-Per-Unit-Effort from commercial fleets and research
15 surveys) are either too coarse-grained or collected over insufficiently large areas for this
16 specific purpose. Various dispersal processes as well as demographic and oceanographic
17 stochasticity are embedded in the incomplete space/time snapshots derived from survey or
18 fisheries data (Tyre, et al. 2001). In an analysis of fishery-independent data from aerial
19 surveys and high-resolution remote-sensing, Royer et al. (2004) stressed out the difficulty of
20 inferring ecological processes from static observed patterns or occupancy data.

21 Archival tags (ATs) and Pop-up Satellite Archival Tags (PSATs) offer a more direct
22 way to study fish behaviour, through the combined records of light (to estimate geolocation),
23 depth and ambient temperature (see Block and Stevens 2001 for a review). While displaying
24 significant drawbacks (e.g. price per unit, premature detachment, satellite transmission
25 failures), this technology has provided valuable insights in the extension and duration of
26 transatlantic migration of BFT, as well as indications about depth and thermal preferences
27 (Block, et al. 2001, De Metrio, et al. 2002, Lutcavage, et al. 1999). Still, these recent results
28 remain largely descriptive and are inferred from a relatively small number of tags. Moreover,
29 individual trajectory is indirectly derived from light-based positions, and is often incomplete
30 or impaired by large observation errors (see e.g. Sibert and Fournier 2001). Challenging tasks
31 are: (1) to reduce uncertainties in locating a tag from light records and other variables, (2) to
32 define a statistical framework taking in account location errors, process errors of movement
33 models and partially observed oceanic variability, to derive useful estimates of global and

1 local (time-varying) movement parameters, (3) to develop protocols to link observed patterns
2 with ecological processes of interest, and (4) to investigate the question of optimally designed
3 tagging strategies. We show here how a general class of Bayesian filters (Evensen 1994,
4 Jonsen, et al. 2003, Morales, et al. 2004) can maximize the output of PSAT data analysis,
5 using non-linear and possibly non-stationary motion models. This approach provides a
6 general framework for jointly exploiting light, environmental and behavioural information
7 from archival tags. First, the theoretical basis of such state-space models is detailed. An
8 example based on simulated PSAT data is then given to assess the performance of the
9 proposed Bayesian filter, using sea surface temperature fields from an Ocean General
10 Circulation Model. We show how such methods can potentially cope with some critical
11 problems induced by archival tagging, conditionally on some specific preliminary work. We
12 seek in particular the inference of high-level information from low-level positional data, to
13 use the track as a random sample of the habitat of the animal.

14

15 **METHODS**

16 **1. State-space modelling**

17 *1.1 Generalities*

18 Data assimilation aims at optimally combining information from imperfect models and
19 imperfect measurements (Tarantola and Valette 1984). It provides a sound approach to
20 location estimation when only low-level sensors and inaccurate representations of the
21 environment are available. Data assimilation can be formulated in a Bayesian context, for
22 example using the following parametric state space model:

$$23 \quad X_{t+1} = f(X_t, \theta_t) + v_t \quad (1)$$

$$24 \quad Y_t = g(X_t, \theta_t) + e_t \quad (2)$$

25 where X_t is the model state vector including animal location, $f(X_t, \theta_t)$ is a function describing
26 the fish state dynamics, $g(X_t, \theta_t)$ is an observation function, and v_t , e_t are process and
27 measurement errors. θ_t is an unknown parameter vector, potentially dependant on the
28 environment. In the Bayesian framework, solving this problem is equal to estimating a
29 posterior probability density $p(X_t | Y_{1:t})$, where $Y_{1:t} = \{Y_1, \dots, Y_t\}$ is the historical observation
30 vector. This is done using a two-step algorithm, or filter. A prediction step (eq. 3) is first

1 performed by computing $p(X_t|Y_{1:t-1})$ knowing $p(X_{t-1}|Y_{1:t-1})$ and a transition probability
2 $p(X_t|X_{t-1})$:

$$3 \quad p(X_t|Y_{1:t-1}) = \int p(X_t|X_{t-1}) p(X_{t-1}|Y_{1:t-1}) dX_{t-1} \quad (3)$$

4 The Bayes rule (eq. 4) is then used to update $p(X_t|Y_t)$ with the newly available observation
5 vector $Y_{1:t}$. This is known as the correction step:

$$6 \quad p(X_t|Y_{1:t}) = \frac{p(Y_t|X_t)p(X_t|Y_{1:t-1})}{p(Y_t|Y_{1:t-1})} \quad (4)$$

7 The density $p(X_t|Y_{1:t})$ then provides an estimate of the hidden state X . In the following part,
8 we give a short presentation of two implementations of the Bayes filter in the linear and
9 general case, as well as their applicability to the geolocation problem in archival tagging.

10

11 *1.2 The Standard Kalman Filter*

12 The Kalman Filter (or KF, Kalman 1960) provides an analytical solution to the state-
13 estimation of order-1 autoregressive processes, i.e. $p(X_t|Y_{1:t-1})$ of Eq. 3 is estimated
14 sequentially and replaced by $p(X_t|Y_{1:t})$. This filter is the best linear unbiased estimator, and is
15 optimal if the dynamics and the observation process are strictly linear, with normally
16 distributed errors (the formulation of the KF is detailed in Appendix 1). Developed in the
17 signal processing community, the KF has been used recently in fish stock models (Schnute
18 1994), catch-at-length data analysis (Sullivan 1992) or movement modelling in salmon and
19 bigeye tuna (Newman 1993, Sibert, et al. 2003). This approach has led to valuable results for
20 the objective analysis of archival tags data, and its main advantage are its ease of
21 implementation in the linear case, and its ability to yield macro-statistics (e.g. the diffusion
22 coefficient in the normal diffusion case).

23 However, the underlying model of the random walk is based on the diffusion equation
24 which, although flexible and tractable, is difficult to interpret and relies on some unlikely
25 assumptions with little ecological meaning (Turchin 1998). Moreover, as noted by Sibert et al.
26 (2003), the KF suffers from drawbacks in the non-linear and non-stationary case, a common
27 situation in animal ecology when time-varying behaviours are considered (e.g. foraging
28 versus migration). Alternative models are also needed to describe super- or sub-diffusion
29 patterns, as seen around aggregating features or during trans-oceanic migration. Data on
30 ambient temperature need also to be used as additional sources of information (Sibert et al.

1 2003), thus inducing strong non-gaussianity in the observation densities. Non-gaussianity may
 2 also occur in the vicinity of shorelines, since probability distributions must be truncated to
 3 exclude position estimates on land. Closed-form analysis of such dynamical systems cannot
 4 be achieved using a standard KF. Still, sub-optimal analysis may be performed using the
 5 Extended Kalman Filter (EKF) or its recent alternatives, the Ensemble Kalman Filter (EnKF)
 6 and the Particle Filter (PF).

7

8 *1.3 The Monte Carlo Particle Filter*

9 Recently, a more general class of Bayesian filters has been proposed and used
 10 successfully to sequentially update the posterior distribution using Monte Carlo methods
 11 (Doucet and Godsill 1998). These methods (collectively called Particle Filters or PF)
 12 approximate the predicted distribution by a set of Dirac delta functions $\delta(\cdot)$, referred to as an
 13 ensemble of particles $X_t^{(i)}$ with corresponding importance weights $w_t^{(i)}$ (Eq. 5).

$$14 \quad p(X_t|Y_t) \approx 1/N \sum_1^N w_t^{(i)} \delta(X_t - X_t^{(i)}), \quad i=\{1, \dots, N\} \quad (12)$$

15 No explicit assumption about the form of $p(X_t|Y_t)$ is made: the PF can be applied to general
 16 nonlinear, non-gaussian systems. These non-parametric techniques can be highly efficient in
 17 that they allow to focus computing power in regions of high likelihood in the state space.
 18 Non-stationary models can also be studied in a straightforward way by extending X with a
 19 slowly (compared to the dynamics) time-varying parameter vector. For example, geolocation
 20 or temperature measurement bias can be modelled as a weak random walk with fixed or $1/t$
 21 decaying noise. While the PF is relatively easy to implement, some improvements must be
 22 considered, such as prior boosting and resampling strategies to avoid filter degeneracy
 23 (Andrieu, et al. 2002).

24 We provide here an algorithm of the generic Particle Filter which can be applied to
 25 any estimation problem (see Doucet and de Freitas, 2001, for more details). For illustration
 26 purposes, an iteration of the PF is presented in Fig. 1.

- 27 • **Givens:** the initial requirements are a dynamical model (Eqs. 1 and 2) with known
 28 state dynamics and observation functions $f(x_t, \theta_t)$ and $g(x_t, \theta_t)$, initial prior densities for
 29 the state vector $p(X_0)$ and parameter vector $p(\theta_0)$, and noise densities $p_{v,t}$ and $p_{r,t}$. The
 30 filter parameters are the number N of simulated particles and the roughening noise

1 density functions $p_{r,t,x}$ and $p_{r,t,\theta}$ (these are time dependent instrumental noises for
 2 smoothing and identification purposes).

3 • **Initialisation step:** Generate N samples, or particles, from the initial state density
 4 $X_0^{(i)} \sim p(X_0)$, with $i=\{1, \dots, N\}$. Each particle is given a weight $w_0^{(i)}:=1/N$.

5 • **Density propagation:** for each particle $i=\{1, \dots, N\}$, we sample the state evolution
 6 noise $v_t^{X,(i)} \sim p_{v,X}$, the parameter evolution noise $v_t^{\theta,(i)} \sim p_{v,\theta}$, and the associated
 7 instrumental noise $r_t^{X,(i)} \sim p_{r,X}$, $r_t^{\theta,(i)} \sim p_{r,\theta}$. The new state and parameter vector are
 8 simulated for each particle using Eqs. (13) and (14):

$$9 \quad X_{t+1|t}^{(i)} = f(X_t^{(i)}, \theta^{(i)}) + v_t^{X,(i)} + r_t^{X,(i)} \quad (13)$$

$$10 \quad \theta_{t+1|t}^{(i)} = \theta_t^{(i)} + v_t^{\theta,(i)} + r_t^{\theta,(i)} \quad (14)$$

11 • **Assimilation step:** the measurement update is achieved by first updating the
 12 weights with a likelihood function for each particle $i=\{1, \dots, N\}$:

$$13 \quad w_t^{(i)} = w_{t-1}^{(i)} p(Y_t | X_t^{(i)}) \quad (15)$$

$$14 \quad w_t^{(i)} = w_{t-1}^{(i)} p_{e,t}(Y_t - g(x_t^{(i)}, \theta_t^{(i)})) \quad (16)$$

$$15 \quad w_t^{(i)} = w_t^{(i)} / \sum_i w_t^{(i)} \quad (17)$$

16 where Eq. (17) corresponds to weight normalization. The posterior distribution is then
 17 obtained by applying Sampling Importance Re-sampling to the weighted particles
 18 (Rubin 1987). This is achieved by sampling with replacement N particles from
 19 $\{X_t^{(i)}(\theta_t^{(i)}, w_t^{(i)})\}$ with probability $w_t^{(i)}$. All samples are then given uniform weights
 20 $w_t^{(i)}=1/N$. The distribution of particles at time t now provides an approximation of
 21 $p(X_t | Y_t)$ as in Eq. (12). The whole process is then reiterated for time step $t + 1$. The
 22 PF thus constructs a sequential estimate of $X_{1:t}$, as in the KF.

24 2. Application to fish tracking: a simulated case study

25 2.1 Introduction

26 In contrast, Morales, et al. (2004) used a recursive Monte-Carlo technique to fit
 27 multiple random walks on Elk movement data, using the WinBUGS package (Spiegelhalter,
 28 et al. 1996). The purpose of this software is to ease the use of Markov Chain Monte-Carlo
 29 methods in statistical analysis. We chose not to use it here since it is mostly designed to
 30

1 handle parametric density functions which are not available in our case (e.g. ambient
2 temperature), while convergence may become an issue for very large or poorly observed data
3 sets. We therefore implemented a sequential PF using the MATLAB 6.5 software. In our
4 PSAT estimation problem, the state vector is $X_t = \{\text{longitude}_t, \text{latitude}_t, \text{depth}_t, \text{temperature}_t\}$
5 for which observations are available directly (from on-board depth and thermal sensors) or
6 indirectly (through light-base geolocation). Now that the state and observation spaces are
7 specified, a model of the fish dynamics in its environment is needed. We restricted this
8 analysis to a two-dimensional space (the third dimension could be included in another model,
9 as the vertical behaviour of pelagic fishes relies on highly different mechanisms).

10

11 *2.2 Definition of the process of fish movement*

12 Quantifying the dispersal of organisms is a key step in analysing the distribution and
13 fluctuations of animal populations (Turchin 1991). Thus, a wealth of models has been
14 developed for simulating randomly moving particles on a plane. Purely statistical models such
15 as the classical Random Walk, Fractional Brownian Motion or the Levy Flight (Viswanathan,
16 et al. 2001) can be distinguished from purely mechanistic models. Among them, the
17 speed/angle based model (Bovet and Benhamou 1988, Wu, et al. 2000) is getting increasingly
18 popular, and has found applications in terrestrial ecology (Morales, et al. 2004), avian studies
19 (Thorup, et al. 2000) and marine ecology (Hubbard, et al. 2004). While not fundamentally
20 different from a diffusion process at large time scales, this framework becomes more relevant
21 at local scales, yielding easier to interpret statistics. Speed and turning angle are also
22 differentiable state variables, whereas the drift and diffusion in the classical Random Walk
23 can not be separated locally. By allowing speed/angle correlations to vary between successive
24 time steps, one can generate variable movement patterns according to the migration/foraging
25 paradigm (Ramos-Fernandez, et al. 2003, Atkinson, et al. 2002, Bartumeus, et al. 2002,
26 Mårell, et al. 2002, Viswanathan, et al. 1999, Krakauer and Rodriguez-Girones 1995). Such
27 dual motion model has recently been documented for pelagic fishes, who move in a relatively
28 featureless environment with no long-range visibility (Newlands, et al. 2004). Switching
29 between a "foraging" behaviour and a "migration" behaviour may then be driven by a
30 combination of external variables (e.g. light, prey density) and internal state (e.g. reproduction
31 period). We generated the trajectory of a randomly moving bluefin tuna using the following
32 hierarchical statistical process:

1 • **Movement mode:** the nature (i.e. forage or migrate) of the movement regime at
2 time t is modelled as an arbitrary binomial process with probability α_t :

$$3 \text{ Mode}_t \sim B(\alpha_t, 1-\alpha_t) \quad (11)$$

4 • **Speed and azimuth:** for each step t , the individual's speed is sampled from a
5 Weibull density of mean λ_{mode} in $(\lambda_{\text{forage}}, \lambda_{\text{migrate}})$ and scale h_{mode} in $(h_{\text{forage}}, h_{\text{migrate}})$,
6 following the observations of Lutcavage, et al. (2000) on bluefin tuna using ultrasonic
7 telemetry:

$$8 \text{ Speed}_t \sim \text{Weibull}(\lambda_{\text{mode}}, h_{\text{mode}}) \quad (12)$$

9 Azimuth deviation between two consecutive steps is sampled from a wrapped Von
10 Mises density with mean 0 and concentration parameter $\kappa_{\text{mode}} \in [\kappa_{\text{forage}}, \kappa_{\text{migrate}}]$,
11 (Thorup, et al. 2000).

$$12 \text{ Angle}_t \sim \text{VonMises}(0, \kappa_{\text{mode}}) \quad (13)$$

13 Intuitively, one can expect migration movements to be more directed and thus, display
14 stronger time correlation, while foraging movements may be more scattered yielding higher
15 residency times. In the stationary case (i.e. constant speed distribution), it is possible to relate
16 such a model to the standard diffusion by computing the expected square displacement using
17 the formulae proposed by Wu, et al. (2000).

18 *2.3 Synthetic data generation and scenarios testing*

19 To illustrate the proposed data assimilation method for analysing archival tags, we
20 assimilated ambient temperature in a two-dimensional case study.

21 • **Simulated track :** Movement was generated over 240 days between the Gibraltar
22 strait and the mid-Atlantic, with a random alternation of directional and adirectional
23 motion regimes (Fig. 2). Time step duration was set to one hour, consistently with the
24 recording resolution in pop-up archival tagging. A geolocation process was then
25 simulated by subsampling the trajectory at a 24-hour period and adding gaussian noise
26 (error sources relate to the time resolution of the recordings, the vertical behaviour of
27 the fish, cloudiness, the proximity of equinoxes, and the optical characteristics of the
28 upper layers (Musyl, et al. 2001 and Welch and Eveson 1999)). Biases were also
29 added to account for possible biofouling or alteration of the light sensors thus inducing
30

1 drifts in the geolocation process. Finally, records of ambient temperature were
2 generated using the sea surface temperature of a weekly output of Mercator PSY-2, an
3 eddy-resolving resolution Ocean General Circulation Model (Baharel, et al. 2002).
4 The temperature fields were smoothed with low-pass filters of various bandwidths, to
5 simulate typical uncertainties and coarse resolution in oceanographic data.

6 • **Scenarios** : The outputs of the Particle Filter assimilating SST were compared to
7 the standard Kalman filter (Sibert and Nielsen 2004) under different noise and bias
8 cases. Issues in latitude estimation were especially investigated by adding gaussian
9 errors ranging from 1 to 3 degrees, while error in estimated longitude were kept low
10 (from 0.5 to 1 degree), a common feature observed in PSATs. Biases in light level
11 sensing/latitude geolocation were simulated as linear drifts (e.g. to account for sensor
12 biofouling) or abrupt shifts (e.g. to account for deterioration) from 1 to 5 degrees in
13 latitude. The SST fields used for temperature assimilation were smoothed to obtain
14 Root-Mean Square Errors (RMSEs) ranging from 0 to 0.5 Celsius degrees, which is
15 the magnitude of the average error in most AVHRR-derived SST products. As the true
16 state of the system was always known, the performance of each method was assessed
17 by computing the RMSE between the estimated state (i.e. the average of the estimated
18 density) and the true (hidden) state time-series.

19

20 RESULTS

21 1. Comparison of the KF and the PF

22 As a verification, we first applied the KF and the PF to state vectors of comparable
23 dimension, i.e. without SST assimilation. Similar random walk models were used for the
24 propagation dynamics, i.e. an advection-diffusion scheme for the KF and its discrete
25 ensemble counterpart for the PF, while all observation errors were held gaussian. Both filters
26 yielded similar results in terms of hidden track estimate and RMSE reduction (results not
27 shown).

28 We then applied the PF and the KF to unbiased series of observations with different
29 errors in the longitude and latitude components, and given SST statistics for the PF. Two
30 examples with low and high geolocation errors are provided as an illustration in Fig. 3. The
31 observation error was set to 0.5 and 1 degree in longitude and latitude for the low deviation
32 scheme, and 1 and 3 for the large deviation scheme. SST error was set to a very low value

1 (0.01°C), which stands for almost perfect *a posteriori* knowledge of the SST field (errors in
2 the forcing SST field are investigated further). Both the KF and PF solutions allowed for a
3 significant decrease in the RMSE of the estimated trajectory compared to observed positions.
4 For weak observation errors, the decrease in the estimate RMSE reached 38% and 50% for
5 the KF and PF, respectively. While the PF yielded a better solution, there were little
6 differences between the KF and the PF estimated tracks (Fig. 3a). The standard Kalman Filter
7 was also outperformed by the Particle Filter for greater observation errors: the observation
8 RMSE was decreased by 42% and 58%, respectively. On average, both filters performed well
9 in this case but qualitatively, the PF estimate still followed accurately the true track, while the
10 KF yielded a smoother estimate (i.e. the KF solution missed a number of local features in the
11 original track –see Fig. 3b). For both cases, the PF solution for the temperature remained
12 almost identical to the true record, as shown in the inserts of Fig. 3. Such result is logical,
13 since a very high confidence was given to temperature data at the assimilation step. Generally,
14 the PF was applied here to a three-dimensional system, thus benefiting from a greater amount
15 of data for its task.

16

17 **2. Assimilating SST**

18 We then investigated to which extent the observation errors were reduced when
19 assimilating SST of varying accuracy. The performance of the PF was expressed as a function
20 of the RMSE of the observed positions as well as the RMSE of the smoothed SST input.
21 About 1000 trials were run with observation errors ranging from 0.5 to 3 degrees and SST
22 error from 0 to 0.5 °C. The propagation of input uncertainty was represented as a response
23 surface (Fig. 4). The observed positions RMSE ranged from 0.8 to 3 degrees, while the PF
24 solution (i.e. estimated track) displayed errors ranging from 0.6 to 0.95 degrees, so that
25 reduction in RMSE attained 25% to 68%. Error on the SST field (in the range studied) had
26 globally little effect on the RMSE of the PF solution: for a given geolocation RMSE, an SST
27 error ranging from 0 to 0.5°C yielded a variation of about 0.1 degree in the final geolocation
28 error. For low geolocation errors (i.e. less than 2 degrees), an increasing SST error increased
29 only slightly the RMSE of the PF solution. For geolocation errors greater than 2 degrees, the
30 opposite situation occurred: low SST RMSE (i.e. less than 0.3 °C) lead to higher error in the
31 PF solution than for high SST RMSE (i.e. more than 0.3 °C). In other words, high confidence
32 in the knowledge of the temperature field degraded the performance of the filter when
33 combined with noisy geolocations, while less accurate (i.e. smoother) fields lead to better

1 results in this case. As a conclusion, SST error had relatively little effect on the PF estimated
2 track, whereas geolocation error (in longitude, latitude) had the greatest influence. This
3 response surface was obtained with informative prior distributions on both geolocation error
4 and individual motion: misleading priors naturally decreased the performance of the PF.

5

6 **3. Estimating geolocation biases**

7 The PF clearly outperformed the KF when supplied with biased geolocations: three
8 examples are given assuming a fixed latitude bias of 2 degrees (Fig 5a), a linear increase from
9 0 to 5 degrees (Fig 5b), and an abrupt shift at mid-track from 0 to 5 degrees (Fig 5c). These
10 bias schemes were supposed to be unpredictable, i.e. relying on unknown causes such as
11 biofouling or sensor alteration. The KF performed well for a fixed bias, since the tagging
12 location and jettison point provide useful information for its correction (Fig. 5a). However
13 this information was not well propagated in time for a linear drift or an abrupt shift (Fig. 5b,
14 4c), yielding to globally poor estimates of the true tracks. When supplied with unbiased
15 temperature data, the PF provided satisfying solutions for biases in observed latitude. For
16 example, a constant bias was identified with $1/t$ time decaying noise, providing a converging
17 solution (Fig 5a). Time-varying biases were identified with a constant noise, yielding rougher
18 solutions (Figs 5b and 5c), to allow for drift identification and shifts detection. Assimilating
19 external temperature apparently gave a decisive advantage over the standard KF for bias
20 estimation, especially at mid-track when tagging location or jettison point were distant in
21 time.

22

23 **4. Inferring movement patterns**

24 Using three observable components of the system (i.e. longitude, latitude and ambient
25 temperature), the PF performed well in correcting the observation process, by 1) increasing
26 the signal-to-noise ratio, and 2) estimating a non-observed variable such as the latitude bias.
27 We then assessed its performance in retrieving hidden variables of the dynamical model itself:
28 Fig. 6 displays a typical PF solution for a true track observed with large latitude errors (> 2
29 degrees), with the estimated and true movement regime time series (insert in Fig. 6). The
30 alternation of directional and adirectional movement was globally not well retrieved (the
31 second part of the true track being totally directional). Still, this output illustrates a number of
32 false detections, e.g. segments of the fitted trajectory are identified as originating from

1 directional motion whereas the corresponding true segment was adirectional. Such result can
2 be observed in segments of high sinuosity, where a large number of adirectional detections
3 occurred. The identification of such hidden variable was highly sensitive to initial priors and
4 ensemble size. Percentages of good identifications reached 70% for small geolocation errors,
5 while it was not significantly different from 50% for higher geolocation errors (e.g. more than
6 2 degrees). Thus, for high observation errors, the PF did not yield better results than a blind
7 guess in the retrieving the hidden variables of the dynamical system.

8

9 **DISCUSSION**

10 The proposed Monte Carlo filter performed well in filtering the track of an individual
11 in a 2-D geographical space (i.e. longitude and latitude). When applied to a linear random
12 walk and gaussian errors, the Particle Filter with Sampling-Importance-Resampling step was
13 comparable to the Kalman Filter. This validated the approximation scheme and the
14 representation of probability distributions by an ensemble of particles. The PF was then
15 extended to a higher observation space (i.e. longitude, latitude and temperature), and a more
16 complex random walk model with regime shifts. This induced strong non-gaussianity in the
17 state distributions, particularly in the latitudinal dimension (see examples in Fig. 7).
18 Multimodal and highly skewed distributions were generated by sea surface temperature, a
19 feature that could be handled by the non-parametric nature of the PF.

20 Including SST as a supplementary variable in the space-time analysis allowed a higher
21 RMSE reduction than in the standard Kalman filter. For high observation errors, the KF
22 behaved like a smoother since geolocation data contributes less to the variability of the final
23 solution. Temperature assimilation prevented excessive relaxing in the PF estimate, thus
24 giving insight in the true variability of the track. Obviously, the PF handled a greater amount
25 of information, since it was applied to a space of higher dimension (i.e. longitude, latitude and
26 temperature). The PF's ability to handle a variety of parametric/non-parametric state and
27 parameters distributions is one of its key features. However, its performance is conditioned by
28 two major issues. First, external forcing variables such as the temperature field need to be
29 spatially informative, and second, the statistics of the observation errors need to be well
30 known for the PF to perform correctly. We illustrated this by using SST fields smoothed with
31 gaussian filters of varying bandwidths. This allowed to preserve their spatial pattern: thus, the
32 temperature RMSE had little effect on average on the RMSE of the PF solution. The

1 Pathfinder AVHRR-derived SST, at its current status, presents a bias/error scheme of
2 $0.00\pm 0.24^{\circ}\text{C}$, which is more than adequate for such analysis (Keogh, et al. 1999).
3 Interestingly, the combination of high observation error and high precision in temperature
4 generates a high error in the PF estimate. This is interpreted as a greater spatial
5 indetermination of the system since sea surface temperature may display pseudo-periodic
6 properties, thus generating multiple choices. As a general result, geolocation uncertainty may
7 cover areas where several water masses display a suitable temperature distribution, thus
8 generating multimodal densities (see Fig. 7). The probability density functions of these state
9 estimates were retrieved by applying gaussian kernels to the updated ensembles, thus
10 allowing non-parametric representations of the probable dispersion in latitude. Issues about
11 the SST spatial structure could be explored *a priori* by comparing geolocation errors (in both
12 latitude and longitude) with empirical variograms or spatial autocorrelation of the SST field.

13 On-line estimation of hidden variables was also performed, both in the observation
14 and state space. Retrieving time-varying bias in latitude, for example, was achieved by
15 allowing the model parameters to change with time as a random walk. A fixed bias was well
16 identified by using a $1/t$ time decaying noise, thus providing converging estimates. Shifts in
17 the bias were identified using randomly walking parameters (i.e. with noise constant in time):
18 this allowed to detect changes in bias, at the cost of a rougher estimate. Regime shifts
19 identification therefore relies on the specification of a suitable noise model in hidden
20 variables. Such estimation is achieved using information on the tag's initial and final positions
21 and the temperature field (it is therefore critical that prior SST fields are free of strong bias for
22 this task). Their accuracy must be assessed as a preliminary work, since it is difficult to
23 estimate both geolocation errors and modelled SST bias at the same time in state-space
24 models. Regime shifts in the dynamic process itself (i.e., migration vs foraging behaviour)
25 were also estimated, yielding less satisfying results: a number of false detections occurred, in
26 relation to both model specification (e.g. autocorrelation of regimes in time) and geolocation
27 errors. This was related to a greater indetermination of the system, i.e. several combinations
28 of parameters have equal probabilities to yield the observed pattern in the data. In this case,
29 the PF solution was not significantly better than a blind guess of the movement regime (i.e.
30 equal probability of choosing between a "directional" and an "adirectional" step). Moreover,
31 the PF appeared to favour the "adirectional" movement, since this least restrictive regime may
32 have yielded higher likelihoods. This feature can be directly related to inverse diffusion, a
33 typically ill-defined problem in which a given state/observation can originate from multiple

1 sets of parameters or initial states. Overcoming this problem depends on reducing the degree
2 of freedom of the system, i.e. through a better formulation of system noise and hidden
3 processes. Supplemental knowledge could be added, as derived from direct observation, such
4 as swimming speed distributions in feeding grounds, frontal zones, or during migration
5 periods. This would also allow to include testable hypothesis in the analysis of archival tag
6 data, by linking the behaviour of the individual to its immediate surroundings. The PF offers a
7 flexible tool for estimating the trajectory of individuals with an assimilation of data on their
8 physical environment, but it is not without any drawback, since its great flexibility can be a
9 handicap for process identification (i.e. the system can quickly become under-determined
10 when supplied with insufficient data). Such drawback was illustrated by the poor estimation
11 of hidden processes variables (i.e. regime shifts in movement statistics): the combined
12 estimation of states and parameters is not easily performed in a single step in a Bayesian
13 framework. Statistical analysis of the track may be performed as a subsequent step, e.g. the
14 most probable trajectory can then be summarized through standard lagrangian statistics, or
15 used in posterior studies for deriving environmental preferences, or assessing interactions
16 with various oceanic features (e.g. fronts and eddies).

17 Linking fish behaviour to its habitat relies on some optimal knowledge of
18 environmental heterogeneity: we therefore need to describe oceanographic variability over a
19 sufficient range of scales. Interfaces between water masses and transition/frontal zones can be
20 of particular interest, since ecological processes of great importance occur at these locations
21 (Bakun 1996). The proposed lagrangian method would allow to locate a fish in relation to its
22 habitat over periods of days to months, thus providing insights on how top predators with high
23 metabolism can interact with the high patchiness of their environment. Species such as bluefin
24 tuna can be seen both as mobile sampler and integrator of oceanic variability, having
25 developed specific strategies to preserve homeostasis. Thermoregulation, fat storage are such
26 well-known physiological mechanisms. Specific foraging strategies are also believed to
27 maximize expected food intake, through opportunistic feeding or Brownian search. Still, little
28 quantitative information is available on how pelagic predators find food in space: integrated
29 studies linking behavioural knowledge, archival data and oceanographic variability may
30 provide interesting clues about how large fish perceive the ocean landscape. The random walk
31 based on speed/angle variations may allow some insight on how pelagic animals orientate in
32 the pelagic domain. Accurate models of tuna behaviour are especially needed in such
33 analyses: such flexible motion models, although conceptually satisfying and versatile, may

1 not be enough informative, thus forbidding data enrichment and yielding data-driven
2 estimates. Data assimilation theory was elaborated in fields such as applied physics, where
3 processes are well known and described by dynamical equations. Unfortunately, this is not the
4 case in ecology, where the actual dynamics of individuals are most often best described by
5 statistical or empirical models.

6

7 **CONCLUSION**

8 We presented in this paper a general PF algorithm that can be applied to parameter-state
9 estimation in animal behaviour studies. Using an ensemble representation of state
10 distributions and sampling-importance-resampling for prior update, we were able to filter the
11 observed trajectory of an instrumented fish. The flexible, non-parametric nature of the PF also
12 permitted the assimilation of external data of general type (here sea surface temperature), thus
13 allowing integrated bio-physical oceanographic studies. We derived a range of observation
14 scenarios from various sets of simulated errors: this pointed out the need for accurate
15 knowledge of the underlying processes and their variability. Gathering data and formalizing
16 hidden processes is an important preliminary step in Bayesian analysis. Only then may the
17 output of such inverse modelling studies be maximized: it was shown in particular that
18 process identification is conditioned by the reduction of observation errors and a clear
19 formulation of the modelled dynamics. This approach opens the gate to a variety of integrated
20 studies at the crossroads of marine behavioural ecology and physical oceanography.

21

22 **ACKNOWLEDGEMENTS**

23 We thank Olivier Maury who provided helpful comments and advice on a earlier version of
24 this manuscript. This research was made possible by a PhD grant funded by CLS and
25 IFREMER for FR.

26

27 **REFERENCES**

- 28 Andrieu, C., Davy, M. and Doucet, A. 2002. Efficient Particle Filtering for Jump Markov
29 Systems. - IEEE transactions on Signal Processing 51: 1762-1769.
30 Atkinson, R. P. D., Rhodes, C. J., Macdonald, D. W. and Anderson, R. M. 2002. Scale-free
31 dynamics in the movement patterns of jackals. - Oikos 98: 134-140.

- 1 Bahurel, P., De Mey, P., Le Provost, C. and Le Traon, P. Y. 2002. A GODAE prototype
2 system with applications: Example of the Mercator system. "En route to GODAE" 2002
3 Proceedings - CNES (ed.) France
- 4 Bakun, A. 1996. Patterns in the Ocean: Ocean Processes and Marine Population Dynamics. -
5 University of California Sea Grant, San Diego, California, USA, in cooperation with
6 Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico.
- 7 Bartumeus, F., Catalan, J., Fulco, U. L., Lyra, M. L. and Viswanathan, G. M. 2002.
8 Optimizing the encounter rate in biological interactions: Lévy versus Brownian
9 strategies. - Physical Review Letters 88.
- 10 Block, B. A., H. D., Blackwell, S. B., Williams, T. D., Prince, E. D., Farwell, C. J., Boustany,
11 A., Teo, S. L. H., Seitz, A., Andreas, W. and Fudge, D. 2001. Migratory Movements,
12 Depth Preferences, and Thermal Biology of Atlantic Bluefin Tuna. - Science 293: 1310-
13 1314.
- 14 Block, B. A. and Stevens, E. D. 2001. Tuna. Physiology, ecology, and evolution. - Academic
15 Press.
- 16 Bovet, P. and Benhamou, S. 1988. Spatial analysis of animal movements using a correlated
17 random walk model. - J. Theor. Biol. 131: 419-433.
- 18 De Metrio, G., Arnold, G. P., Block, B. A., De la Serna, J. M., Deflorio, M., Cataldo, M.,
19 Yannopoulos, C., Megalofonou, P., Beemer, S., Farwell, C. and Seitz, A. 2002.
20 Behaviour of post-spawning atlantic bluefin tuna tagged with pop-up satellite tags in the
21 mediterranean and eastern atlantic. - Col. Vol. Sci. Pap. ICCAT 54: 415-424.
- 22 Doucet, A. and Godsill, S. 1998. On Sequential Monte Carlo Sampling Methods for Bayesian
23 Filtering. - Cambridge University.
- 24 Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model
25 using Monte Carlo methods for forecast error statistics. - J. Geophys. Res. 99: 10143-
26 10162.
- 27 Fromentin, J.-M. 2003. The East Atlantic and Mediterranean Bluefin tuna management:
28 uncertainties and alternatives. - Scientia Marina 67: 51-62.
- 29 Hubbard, S., Babak, P., Sigurdsson, S. T. and Magnusson, K. G. 2004. A model of the
30 formation of fish schools and migrations of fish. - Ecol. Modell. In press.
- 31 ICCAT. 2003. Report of the 2002 Atlantic bluefin tuna stock assessment session. - Col. Vol.
32 Sci. Pap. ICCAT 55: 710-937.
- 33 Jonsen, I. D., Myers, R. A. and Flemming, J. M. 2003. Meta-analysis of animal movement
34 using state-space models. - Ecology 84: 3055-3063.

- 1 Kalman, R. 1960. A New Approach to Linear Filtering and Prediction Problems. -
2 Transactions of the ASME--Journal of Basic Engineering 82: 35-45.
- 3 Keogh, S. J., Robinson, I. S., Donlon, C. J. and Nightingale, T. J. 1999. The accuracy of
4 AVHRR SST determined using shipborne radiometers. - International Journal of
5 Remote Sensing 20: 2871 - 2876.
- 6 Krakauer, D. C. and Rodriguez-Girones, M. A. 1995. Searching and learning in a random
7 environment. - J. Theor. Biol. 177: 417-419.
- 8 Lutcavage, M., Brill, R., Skomal, G. B., Bradford, C. C. and Howey, P. W. 1999. Results of
9 pop-up satellite tagging of spawning size class fish in the Gulf of Maine: do North
10 Atlantic bluefin tuna spawn in the mid-Atlantic? - Can. J. Fish. Aquat. Sci. 56: 173-177.
- 11 Lutcavage, M., Brill, R., Skomal, G. B., Chase, B. C., Goldstein, J. and Tutein, J. 2000.
12 Tracking adult North Atlantic bluefin tuna (*Thunnus Thynnus*) in the northwestern
13 Atlantic using ultrasonic telemetry. - Mar. Biol. 137: 347-358.
- 14 Mårell, A., Ball, J. P. and Hofgaard, A. 2002. Foraging and movement paths of female
15 reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. -
16 Can. J. Zool. 80: 854-865.
- 17 Mather, F. J., Mason, J. M. and Jones, A. C. 1995. Historical document: life history and
18 fisheries of Atlantic bluefin tuna. - NMFS-SEFSC, p. 165 pp.
- 19 Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E. and Fryxell, J. M. 2004. Extracting
20 more out of relocation data: building movement models as mixtures of random walks. -
21 Ecology in press.
- 22 Musyl, M. K., Brill, R., Curran, D. S., Gunn, J. S., Hartog, J. R., Hill, R. D., Welch, D. W.,
23 Eveson, J. P., Boggs, C. H. and Brainard, R. E. 2001. Ability of archival tags to provide
24 estimates of geographical position based on light intensity. - Reviews in Fish Biology
25 and Fisheries In press.
- 26 Newlands, N. K., Lutcavage, M. and Pitcher, T. J. 2004. Analysis of foraging movements of
27 bluefin tuna (*Thunnus thynnus*). Individuals switch between two modes of search
28 behavior. - Population Ecology in press.
- 29 Newman, K. B. 1993. State-Space Modeling of salmon migration and Monte Carlo
30 alternatives to the Kalman filter. - University of Washington.
- 31 Ramos-Fernandez, G., Mateos, J. L., Miramontes, O., Cocho, G., Larralde, H. and Ayala-
32 Orozco. 2003. Lévy walk patterns in the foraging movements of spider monkeys (*Ateles*
33 *geoffroyi*).

- 1 Ravier, C. and Fromentin, J.-M. 2004. Are the long term fluctuations in Atlantic bluefin tuna
2 population related to environmental changes? - *Fisheries Oceanography* 13: 145-160.
- 3 Royer, F., Fromentin, J.-M. and Gaspar, P. 2004. Association between bluefin tuna schools
4 and oceanic features in the western Mediterranean. - *Mar. Ecol. Prog. Ser.* 269: 249-263.
- 5 Rubin, D. B. 1987. A noniterative sampling/importance resampling alternative to the data
6 augmentation algorithm for creating a few imputations when fractions of missing
7 information are modest: the SIR algorithm. Discussion of "The calculation of posterior
8 distributions by data augmentation" by Tanner and Wong. - *Journal of the American*
9 *Statistical Association* 82: 543-546.
- 10 Schnute, J. 1994. A general framework for developing sequential fisheries models. - *Can. J.*
11 *Fish. Aquat. Sci.* 51: 1676-1688.
- 12 Sibert, J. R. and Fournier, D. A. 2001. Possible models for combining tracking data with
13 conventional tagging data. Symposium on tagging and tracking marine fish with
14 electronic devices Proceedings - Honolulu
- 15 Sibert, J. R., Musyl, M. K. and Brill, R. 2003. Horizontal movements of bigeye tuna (*Thunnus*
16 *obesus*) near Hawaii determined by Kalman filter analysis from archival tagging data. -
17 *Fisheries Oceanography* 12: 141-151.
- 18 Sibert, J. R. and Nielsen, A. 2004. The KFtrack package.
- 19 Spiegelhalter, D. J., Thomas, A., Best, N. and Gilks, W. R. 1996. BUGS 0.5, Bayesian
20 inference using Gibbs sampling. - Medical Research Council Biostatistics Unit, Institute
21 of Public Health.
- 22 Sullivan, P. 1992. A Kalman filter approach to catch-at-length analysis. - *Biometrics* 48: 237-
23 257.
- 24 Tarantola, A. and Valette, B. 1984. Inverse problems - Quest for information. - *J. Geophys.*
25 50: 159-170.
- 26 Thorup, K., Rabøl, J. and Madsen, J. J. 2000. Can clock-and-compass explain the distribution
27 of ringing recoveries of pied flycatchers? - *Animal Behaviour* 60: 3-8.
- 28 Turchin, P. 1991. Translating foraging movements in heterogeneous environments into spatial
29 distribution of foragers. - *Ecology* 72: 1253-1266.
- 30 Turchin, P. 1998. Quantitative analysis of movement. - Sinauer.
- 31 Tyre, A. J., Possingham, H. P. and Lindenmayer, D. B. 2001. Inferring process from pattern:
32 can territory occupancy provide information about life history parameters? - *Ecological*
33 *Applications* 11: 1722-1737.

1 Viswanathan, G. M., Afanasyev, V., Buldyrev, S., Havlin, S., da Luz, M. G. E., Raposo, E. P.
2 and Stanley, H. E. 2001. Lévy flights search patterns of biological organisms. - *Physica*
3 *A.* 295: 85-88.

4 Viswanathan, G. M., Buldyrev, S., Havlin, S., da Luz, M. G. E., Raposo, E. P. and Eugene
5 Stanley, H. 1999. Optimizing the success of random searches. - *Nature* 401: 911-914.

6 Welch, D. W. and Eveson, J. P. 1999. An assessment of light-based geo-position estimates
7 from archival tags. - *Can. J. Fish. Aquat. Sci.* 56: 1317-1327.

8 Wu, H., Li, B.-L., Springer, T. A. and Neill, W. H. 2000. Modelling animal movement as a
9 persistent random walk in two dimensions: expected magnitude of net displacement. -
10 *Ecol. Modell.* 132: 115-124.

11
12

1 **Figures**

2 Figure 1. Iteration of a Particle Filter between $t-1$ and t for $N=10$ particles.

3 Figure 2. Synthetic trajectory of a freely-swimming bluefin tuna. The shift between motion
4 regimes (i.e. directional/adirectional) is governed by an arbitrary binomial process.
5 This random walk has been simulated over 240 days with a 1-hour time step. Insert
6 shows the corresponding along-track temperature sampled from a weekly output of the
7 MERCATOR circulation model.

8 Figure 3. Comparison of the standard Kalman Filter estimate (blue line) and the Particle Filter
9 estimate (dark line) for two unbiased observation processes with error of: a) 0.5 degree
10 in longitude and 1 degree in latitude, b) 1 degree in longitude and 3 degrees in latitude.
11 Actual (hidden) track is in red; observations are materialized by grey crosses.

12 Figure 4. Response surface of the Root Mean Square Error (RMSE) of the Particle Filter
13 output, as a function of the initial geolocations RMSE and the temperature field
14 RMSE.

15 Figure 5. Three scenarios of bias estimation in observed latitude using the Particle Filter: a)
16 with a constant bias of 2 degrees, b) with a linearly increasing bias from 0 to 5
17 degrees, and c) with an abrupt change from 0 to 5 degrees at mid-track. The ensemble
18 estimate of the bias (black thick line) is given along with ensemble 95% confidence
19 intervals (grey lines). The dashed line stands for the true (hidden) bias. The maps on
20 the right show the corresponding biased observations (+), with the KF most probable
21 track (grey line) and the PF estimated track (dark line).

22 Figure 6. Regime shift detection along the simulated track, observed with errors of 0.5 degree
23 in longitude and 1 degree in latitude. Black line is the PF estimated track, dashed line
24 is the true (hidden) track. Observed geolocations are symbolized by crosses. Black
25 dots point out segments estimated as “adirectional” by the PF. Insert shows the time
26 series of the true (hidden) sequence of regimes and the PF estimated sequence, with
27 black lines representing adirectional movement.

28 Figure 7. Latitudinal distribution of the PF-estimated daily geolocations (grey curves). Black
29 dots symbolize the daily ensemble means.

30
31
32

1 **Appendix 1. Formulation of the Standard Kalman Filter**

2

3 Given a linear system yielding observations Y of a dynamical state X:

4 $X_t = AX_{t-1} + BU_{t-1} + v_{t-1}$ (5)

5 $Y_t = HX_t + e_t$ (6)

6 Where U_{t-1} is the drift, $v \sim \mathcal{N}(0, Q)$ is a process noise of variance Q, and $e \sim \mathcal{N}(0, R)$ is an
7 observation noise of variance R. As the errors are assumed gaussian, the density of the state
8 vector is fully described by its two first moments, i.e. its mean state X and covariance matrix
9 P. The filter itself consists first of a time update of both X and P (prediction step):

10 $X_{t|t-1} = AX_{t-1} + BU_{t-1}$ (7)

11 $P_{t|t-1} = AP_{t-1}A^T + Q$ (8)

12 The measurement update (or assimilation step) is then performed by computing the Kalman
13 gain K_t and updating both X and P, using the equations derived by (Kalman 1960):

14 $K_t = P_t H^T (H P_t H^T + R)^{-1}$ (9)

15 $X_t = X_{t|t-1} + K(Y_t - HX_{t|t-1})$ (10)

16 $P_t = (I - K_t H) P_{t|t-1}$ (11)

17

18

19

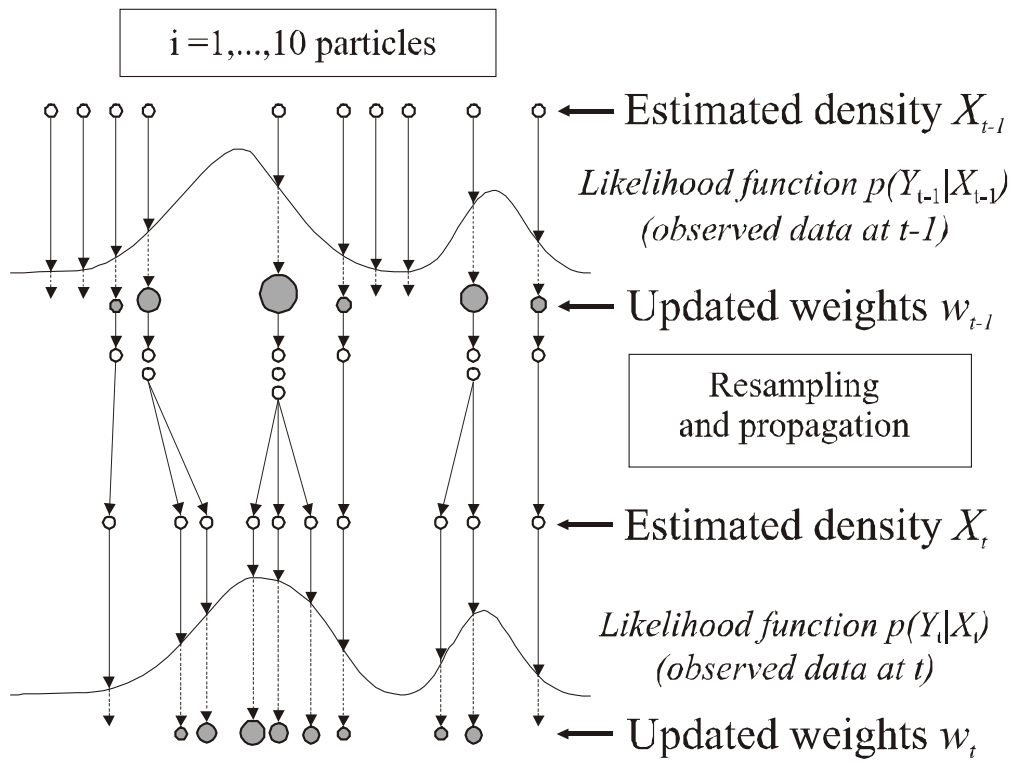


Figure 1

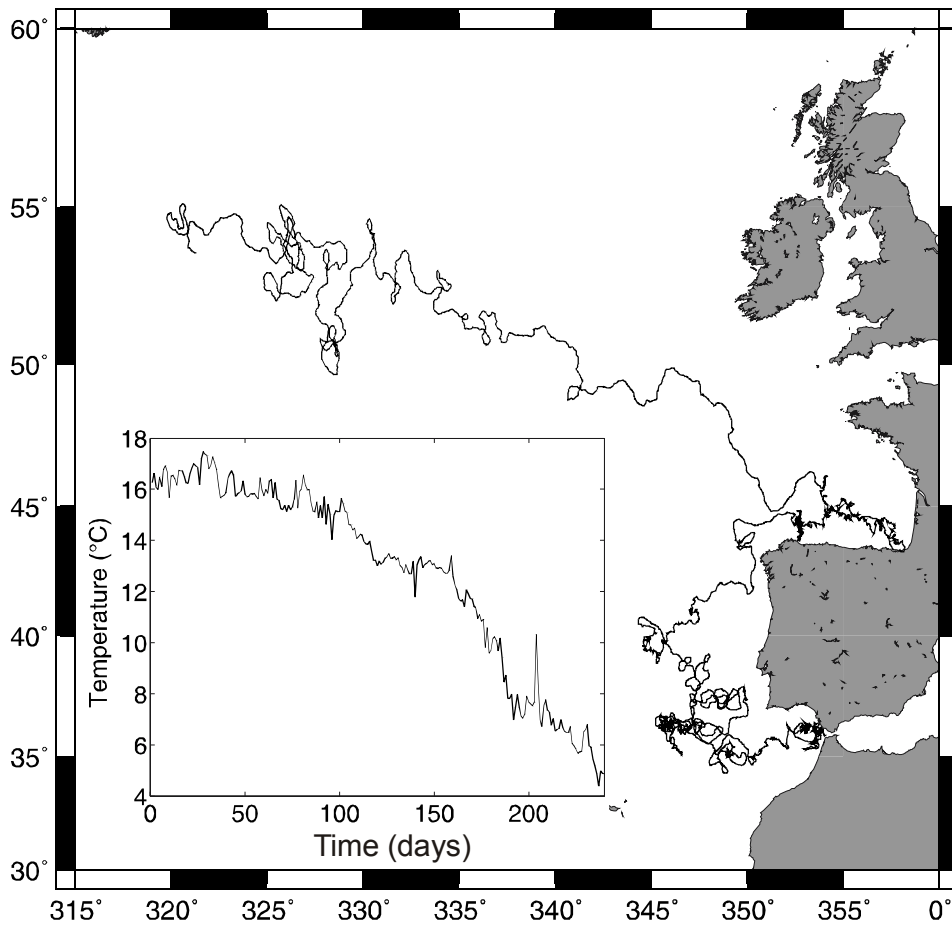


Figure 2

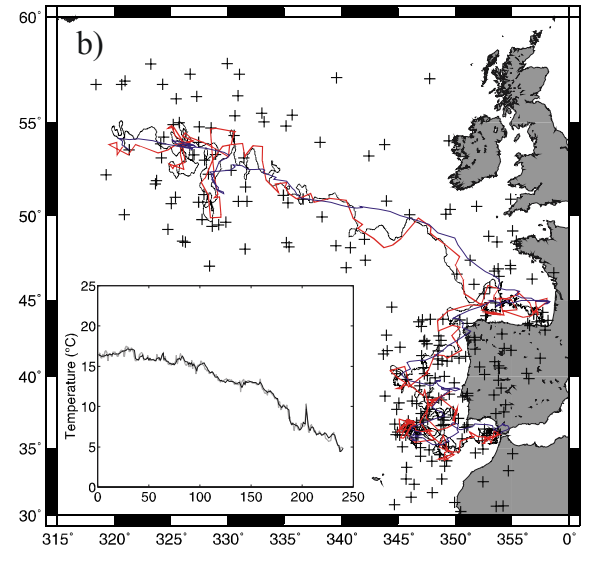
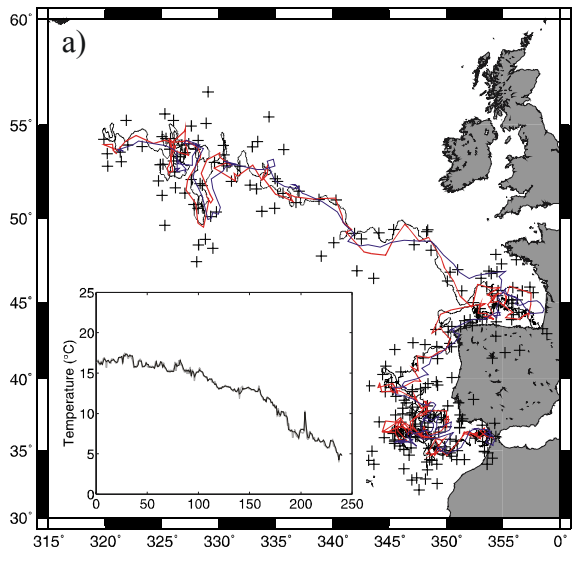


Figure 3

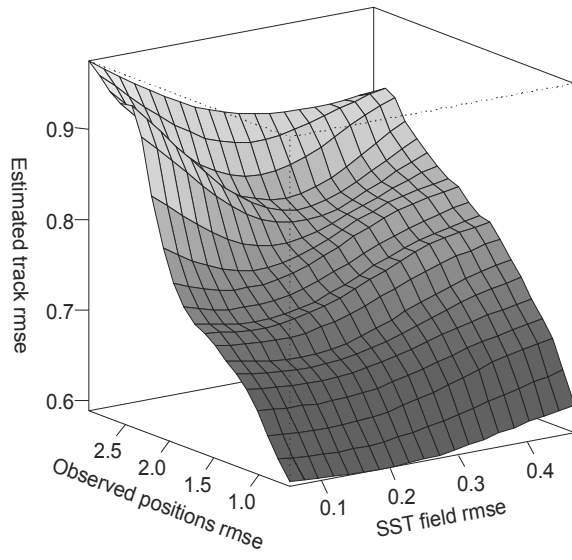


Figure 4

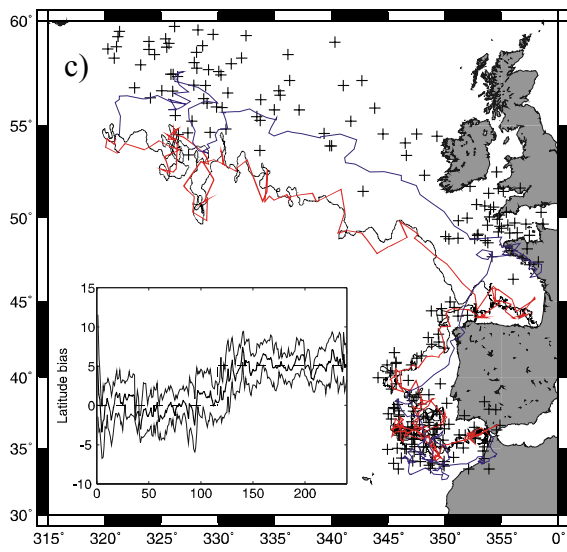
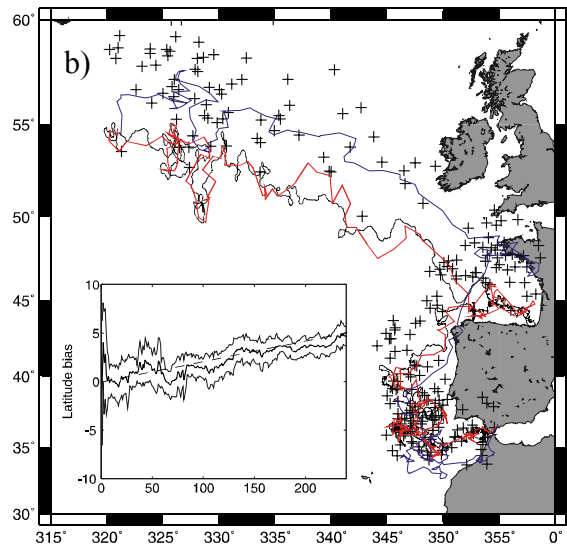
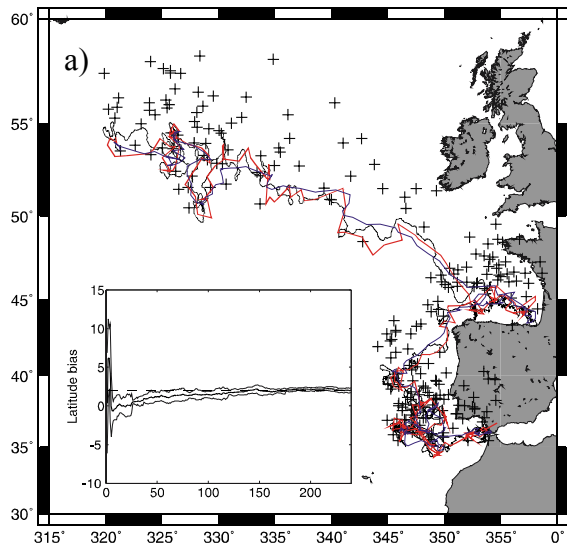


Figure 5

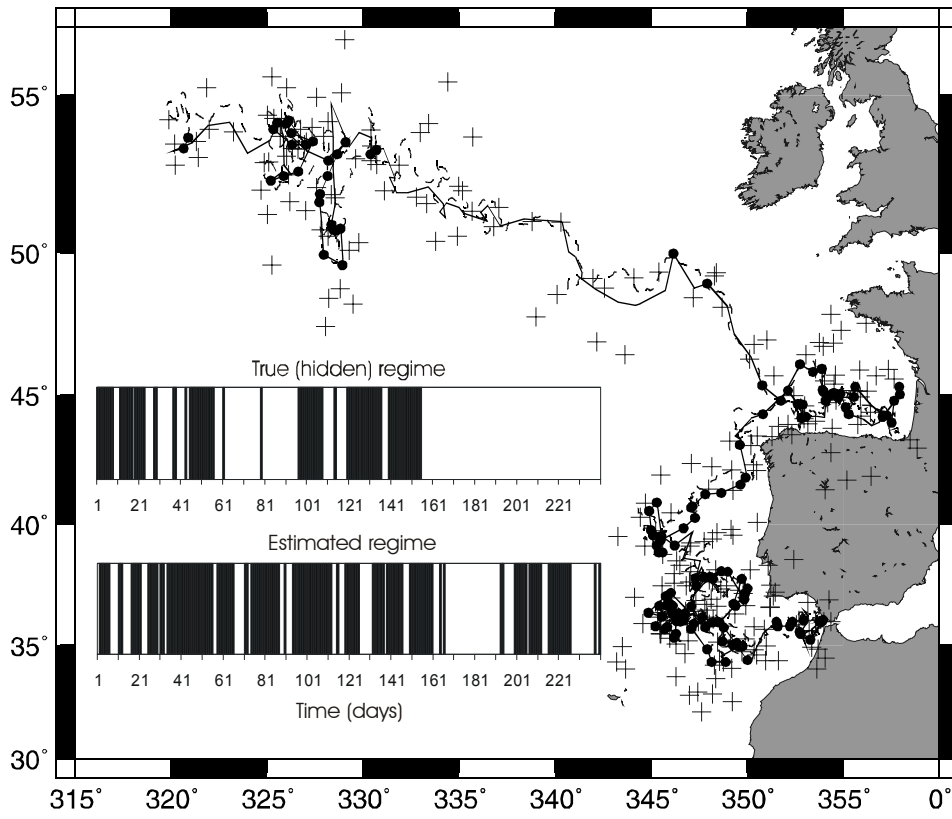


Figure 6

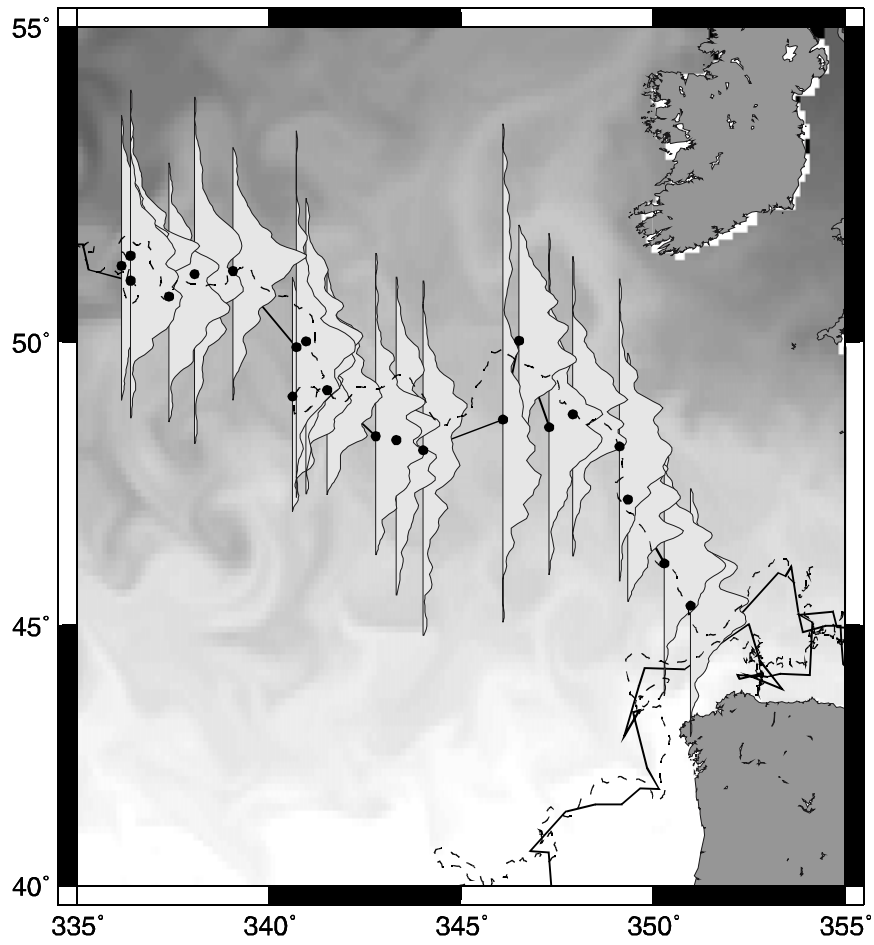


Figure 7

Wayfinding in the Ocean :
Sensors for the Equipment of Archival Tags

A few thoughts...

F. Royer^{1,2}

¹IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean Monnet, BP 171, 34203 Sète cedex, France

²CLS, Division Océanographie Spatiale, 8-10 rue Hermès, 31526 Ramonville St Agne, France

Abstract

This paper hosts a short review of sensors that can potentially be implemented in archival tags. The purpose of these sensors would be to improve geolocation through enhanced navigation and dead reckoning, to provide more detailed insights in fish behaviour, and yield accurate movement statistics suitable for bioenergetic studies. The basic design and theory for these sensors are described, as well as their applicability to archival tagging. In particular, we focus on sensing and recording variations in speed, acceleration and azimuth. Applications such as the detection of hunting or spawning events, or the record of swim speed distributions over long periods of time are presented. Merging this along-track information with actual oceanographic fields would then lead to integrated studies of the behaviour of pelagic predators in their oceanic habitat.

Keywords: Archival tags, pelagic fish, behaviour, accelerometer, compass, polarized light

Introduction

Animal navigation in the open ocean has for long been studied through tag-recapture experiments, short-scale telemetry or sighting reports (Mather 1962; Sibert *et al.* 1999; Lutcavage *et al.* 2000; Bach *et al.* 2003). In the past decade, the archival tag technology has allowed insight in a previously poorly observed field, i.e. fish migration (Gunn *et al.* 1994; Block *et al.* 1998). Accounts of long-range displacement are becoming numerous, yet the existence of precise guidance mechanisms in fish is still subject to debate (Block & Stevens 2001). For many species, it also remains unclear if accurate sources of information can be found in the pelagic realm to allow for absolute and/or relative navigation, or whether the general oceanic circulation combined with simple adaptive behaviour is sufficient to yield the observed migration patterns, thus rendering precise clues unnecessary (Carey 1990). The sensors implemented to date on archival tags provide detailed data on vertical behaviour and ambient or peritoneal temperature, along with limited geolocation ability. How a fish navigates in and makes use of its habitat is still an open issue: improved geolocation is critical for solving this question (Boyd *et al.* 2004; Teo *et al.* 2004; Royer *et al.* 2005). The goal here is 1) to identify sources of data for accurate positioning of archival tags (possibly using "bioinspired" engineering), and 2) to identify sensors that can potentially reveal the processes allowing for animal navigation/wayfinding in the ocean. A short review of such sensors with potential application to archival tags is presented in this paper, eventually allowing to enrich our knowledge of pelagic life. Following some specific needs in archival tag design, we present sensors with few or no moving parts, low power consumption, no light emission and dedicated to motion sensing. We hope this paper will serve as a starting ground for the development of new tag designs and new kinds of field studies.

Measuring speed

Large pelagic predators and tuna-like species are believed to sustain elevated swimming speeds for extended periods of time (Weihs 1973, 1974; Videler & Weihs 1982). However, limited data is available on their Maximum Sustained Swimming Speed, since it involves extended tracking using acoustic telemetry, or experimenting in a water tunnel with captured individuals. Burst swimming speeds are known to attain high values when hunting or escaping predators. So far, such records remain anecdotal and occur in specific, uncontrolled conditions. Sensing and recording the speed of large fish using archival tags may have a wealth of applications, such as the delivery of accurate, long-term statistics on swimming speeds in relation to ecological processes and events, the objective rejection of outliers in

geolocation time series, or even dead reckoning for fine scale analysis of movements in relation to oceanographic features.

A loch-equipped unit would have an external magnetized impeller, tied through a cable to the body of the tag. The sensor itself would be placed in the tag and perform remote measurement of the induced magnetic field. Although this would be the easiest way to measure water speed, this does not appear to be a robust technology: the impeller may act as a vibrating lure and may not sustain the damage caused by neighbouring individuals. The induced magnetic field, although weak, may also alter the functioning of the microchip.

One may propose another technology, based on the Pitot tube, a static measuring device. This approach relies on a conventional sensor (a pressure transducer) for speed sensing, and gets its robustness from the absence of any moving part. This device consists of a probe facing the water flow: several small holes are drilled around the outside of the tube and a centre hole is drilled down the axis of the tube. The outside holes are connected to one side of the pressure transducer. The center hole is kept separated from the outside holes and is connected to the other side of the transducer. The transducer measures the difference in pressure in the two groups of tubes by measuring the strain in a thin element using a piezoelectric gauge. Since the outside holes are perpendicular to the direction of travel, these tubes are pressurized by the local random component of the air velocity. The pressure in these tubes is the static pressure (P_s) discussed in Bernoulli's equation. The center tube, however, is pointed in the direction of travel and is pressurized by both the random and the ordered air velocity. The pressure in this tube is the total pressure (P_t). The pressure transducer measures the difference in total and static pressure. With the difference in pressures measured and knowing the local value of water density ρ , we can use Bernoulli's equation to give us the velocity. Bernoulli's equation states:

$$P_s + (1/2) \rho V^2 = P_t$$

Solving for V gives:

$$V^2 = [2(P_t - P_s)] / \rho$$

More-refined formulae exist, that take account of specific geometry of the probe, the Reynolds number of the flow, and such details. A key point is that the pitot tube measures velocity correctly only when the velocity vector is parallel (or very close to parallel) to its shaft. This is a low tech, yet robust solution: there are no moving parts, and piezoelectric differential transducers may be available at low cost. As the measurement relies on pressure difference, it is independent of the travelling depth or the density of water. No correction or recalibration over time is necessary since the sensor is not subject to drift. The sensitivity should be assessed through basic experiments, but one can expect accurate results since the

high density of water allows high differences between the total and static pressure, largely exceeding the minimum detectable value for miniature sensors. The probe, however, should be conceived with specific materials or treated in regard to fouling: the obturation of the holes may quickly lead to false readings.

Measuring acceleration

Thrust, or acceleration is linked to higher frequency and amplitude in tail beat. Recording acceleration could reveal how fishes of various size and specie can deliver thrust when escaping predators or pursuing preys, or even provide insight in some specific strategies such as swim-and-glide (Videler & Weihs 1982). Tunas may reach burst speeds of up to 10 FL/s, potentially providing intense signals of their feeding activity. Foraging success is usually assessed *a posteriori* from stomach contents (Ménard & Marchal 2003): more details on feeding behaviour and the frequency of hunting events may possibly be inferred using such sensor, and linked to observed oceanographic conditions or other internal variables. A tail-mounted sensor would also allow to link tail beat to swimming speed for further bioenergetic studies, in a similar fashion as in avian studies (Weimerskirch *et al.* 2004).

Basic concept

Accelerometers are inertial measurement devices that convert mechanical motion to an electrical signal. They consist of a piezoelectric crystal and mass normally enclosed in a protective metal case. The signal is proportional to the acceleration using the piezoelectric principle: as the mass applies force to the crystal, it creates a charge proportional to acceleration. The charge output is measured in pico Coulombs per g (pC/g) terms where g is the force of gravity. A charge amplifier then converts the charge output of the crystal to a proportional voltage output in mV/g terms. Minimum detectable acceleration, bandwidth and cut-off frequency are an accelerometer's characteristics. Unlike a velocity pickup, such transducer has no moving parts subject to normal wear. Therefore, the output sensitivity does not require periodic adjustments to correct for wear.

One-way forward accelerometer

Accelerometers sense variation in speed in only one given direction: such sensor could be mounted in the tag body, to sense longitudinal accelerations. Sampling strategies may consist in sampling repeatedly or averaging acceleration over given time intervals, or

recording the occurrence of value over a given threshold (i.e. binary data). The understanding of the fish's physical activity would greatly benefit of such sensor, since it may reveal periods of time where the individual switches between routine and high-speed, anaerobic metabolism, and thus has greater needs for dissolved oxygen: O₂ demand is believed to be a critical factor in tuna metabolism, especially through oxygen debt recovery (Brill & Bushnell 2001). Intuitively, such sensor would be most useful for recording acceleration values and detecting events of interest: it may be hazardous to use it for assessing swimming speeds, since the necessary integration step can lead to very large errors and is sensitive to drift (i.e. slow-downs due to high speed turns or constant forward accelerations may not always be detectable by a miniature longitudinal accelerometer).

Knowing the azimuth

The heading of a tagged fish can be derived from a magnetic compass, provided there are no sources of disturbance in its immediate surroundings. On the long term, knowing the heading of a tagged fish can allow for more accurate positioning in relation to water masses and boundaries (Block *et al.* 2002). For example, navigation through dead reckoning may be performed by sequential integration of both heading and speed time series (Newlands *et al.* 2004). High frequency turns are likely to occur as part of schooling behaviour in tuna-like species, or during hunting events. To avoid aliasing or noisy signals, one may prefer averaging the measurements over given time intervals, instead of subsampling the output of the compasses. Three variations of the digital compass are presented here.

Two-axis magnetometer

The simplest compass instrument contains a two-axis magnetometer (or two single-axis magnetometers) designed to measure earth's horizontal magnetic field (through its X and Y component). Although the easiest compass idea to implement, it suffers greatly in accuracy when it is tilted. A two-axis compass is useful only when accuracy is not important or when the compass will be accurately levelled prior to use. For these reasons, a simple two-axis magnetometer is unlikely to provide reliable results if implemented in an archival tag: tuna-like fishes may constantly tilt or perform rapid and unpredictable movements when schooling.

Gimballed two-axis magnetometer

To overcome the problem of tilt angle, the compass can be placed on gimbals, thus allowing the sensor to swing and remain level. Gimbaling has two major drawbacks: first, accurate readings are not possible until the compass stops swinging. Second, hard-iron calibration is needed to remove magnetic biases induced by components such as the battery. This solution may therefore not be appropriate for archival tagging.

Three-axis magnetometer with tilt sensor

With this compass, the magnetometers are fixed relative to the tag. The compass also includes a tilt sensor to measure the orientation of the compass. The three-axis magnetometer measures the entire 3-D magnetic field: using the tilt measurement, it computes what the magnetic field intensity would be if the compass was level (X and Y in the horizontal plane). Then it can simply compute the heading using $\theta = \arctan(Y/X)$. The hard iron calibration in this case will be three dimensional: we measure (X_0, Y_0, Z_0) and subtract these constants from all future magnetic measurements. The absence of moving parts is a strength of such design, but its power consumption must be assessed and levelled to meet the requirements of battery life in archival tagging.

Underwater navigation by light

Scattering of sunlight in the earth's upper atmosphere leads to partial polarization of light from the blue sky (Waterman 1954). As a result, the blue sky shows a polarization pattern consisting of e-vectors oriented in concentric circles around the sun. The degree of polarization varies from 0 (direct sunlight) to a maximum (under optimal conditions) of about 0.75 or 75% along the great circle of the sky at an angular distance of 90° from the sun (Coulson 1988). Since the e-vector pattern of the sky is linked to the solar position, directional information in the sky is independent of the visibility of the sun, and can also be obtained from a small patch of blue sky. The use of polarized light patterns in animal navigation has been shown both experimentally and in field studies (Edrich & von Heiversen 1976; Wehner 1976; Horvath & Varju 1995; Phillips *et al.* 2001; Parkyn *et al.* 2003). Horvath and Varju (1995) have shown that the underwater polarization of light through Snell's window is very reminiscent of the polarization pattern in the atmosphere. A number of studies has also reported the sensitivity of marine animals to e-vectors: juveniles of salmon may use it as a cue for long-distance navigation during migration (Parkyn *et al.* 2003), while Horvath and Varju

(1995) demonstrated orientation by polarized light in grass shrimp and rainbow trout. Whether the incidence of atmospheric light rays plays a role in animal orientation at dawn and dusk is still unclear (Novales Flamarique & Hawryshyn 1997). The precise design of a polarization-sensitive device is still to be thought of, and may involve much calibration and validation work. Still, it may be an interesting approach for identifying cues available for underwater navigation.

Single-diode polarimeter

A single-lens polarimeter mounted on top of a tag would mimic the so-called "third-eye" or "pineal eye" found in amphibians and some shark or agnath species. Such eye is unable to provide images of the environment but is sensitive to various wavelengths and polarization variations. The mechanisms involved in sensing and analysing the underwater light pattern for navigation are still unclear, but recent research has shown the ability of fish to use such clues as an aid for migration. In a bio-inspired approach, one could think of using a photodiode shed by a polarizing filter to measure the amount of polarized light in the travelling direction. If the polarization filter is rotated beneath a source of polarized light, the measured light intensity shows a sinusoidal behaviour, reaching its maximum if the e-vector direction and the filter orientation are parallel (Kirschfeld 1972):

$$I_1(\varphi) = I_{\min} + (I_{\min} - I_{\max})\sin^2(\varphi - \varphi_1)$$

With φ the e-vector orientation of the polarized light, φ_1 the angle of maximal sensitivity of the sensor, I_n the light intensity measured by the sensor, I_{\max} the maximal light intensity ($\varphi - \varphi_1 = 0^\circ$), I_{\min} minimal light intensity ($\varphi - \varphi_1 = 90^\circ$). However, such basic design is unable to provide the precise direction of the sun since it relies on rotating the sensor or waiting for scanning behaviour of the fish: values from other directions are needed to yield interpretable patterns of e-vectors.

Tri-diode polarimeter

To determine e-vector direction without rotating the sensor, data from three photodiodes have to be integrated simultaneously. Those three sensors would be covered by polarizing filters oriented at angles φ_1 , φ_2 , and φ_3 . From the data I_n of the three sensors, φ can be computed independently of the overall light intensity and degree of polarization using the following equation:

$$\tan(2\varphi) = \frac{(I_2 - I_3)\cos(2\varphi_1) + (-I_1 + I_3)\cos(2\varphi_2) + (I_1 - I_2)\cos(2\varphi_3)}{(-I_2 + I_3)\sin(2\varphi_1) + (I_1 - I_3)\sin(2\varphi_2) + (-I_1 + I_2)\sin(2\varphi_3)} \quad n=1,2,3$$

The e-vector orientation is also a function of the time of the day: post-processing must take in account the daily westward movement of the sun of about 15° per hour. A steady platform is also needed: the sensors must point upward in order to measure a minimal amount of light from the surface. Calibration experiments are needed on moored buoys, at different depths and time of the year to investigate such issue. Few papers report such studies (Tyler 1963; Horvath & Varju 1991).

Perspectives

Tag engineering

Bioinspired robotics and biomimetic sensing technology are areas of active research. Hard engineering problems in archival tags concern power consumption, the overall dimensions of the device, and memory size. Once these issues are resolved, it is time to focus on the sensor type to implement and the precise events to record. For example, recording only "events of interest" may be an interesting way of compressing data. High order information such as rate of change may be less accurate since their integration is subject to errors and drifts but they may be more interesting for monitoring ecological processes/events.

Tagging techniques

Some of the proposed designs would benefit from new tagging techniques, such as harnessing the tag on top of the fish to prevent tilting and to ensure sensors are correctly aligned in regard to the body or the surface.

Data processing and analysis

The idea is to perform geolocation through the integration of dead reckoning, ambient temperature assimilation, light positioning and possibly inertial navigation. An issue to resolve is the question of sampling, averaging or integrating, each of these techniques having specific drawbacks. Data fusion algorithms are also needed to perform dead reckoning or inertial navigation: such algorithms may consist in Kalman-type sequential filters and their non-linear counterparts (Jonsen *et al.* 2003; Sibert *et al.* 2003; Royer *et al.* 2005). Dead reckoning is the process of measuring vehicle attitude (roll, pitch, yaw) and velocity (forward, lateral moves and net water flow) and integrating these measurements over time from an initial position to determine the current position. Attitude and velocity measurement errors will cause the integrated position estimate to deviate from the true position over time. Absolute measurements are therefore needed to prevent unbounded error growth: such

information may be provided by time series of light-derived geolocations, which will serve therefore as a secondary source of data. Dead reckoning is very demanding on sensors, but with a combination of relative and absolute position measurement, one may eventually achieve high precision in tag geolocation.

Conclusion

These thoughts have been partially inspired by some recent work in biomimetic robotics. This is an attempt to forecast possible advances in archival tagging for pelagic species, with a focus on enhancing accuracy in geolocation, and better insights in ecological processes (e.g. feeding in patches or opportunistic foraging, directional or random orientation in the sea, swimming strategies...). Researchers in bioenergetics may greatly benefit from such recorded data: the frequency and cost of long-range swim or foraging may be assessed over long tagging periods. An interesting perspective would be to achieve the highest possible accuracy in geolocating the tag, which could be combined with a suitable description of the animal's immediate surroundings. The variability of pelagic fish behaviour may then be linked to its habitat: process inference and hypothesis testing may then be performed over a range of spatial scales, to understand how it makes use of its environment.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Recurrent and density-dependent patterns in the long-term fluctuations of Atlantic bluefin tuna trap catches

F. Royer^{1,2*}, J.-M. Fromentin¹

¹ IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean Monnet, BP 171, 34203 Sète cedex, France

² Current address for FR. Large Pelagics Research Lab, University of New Hampshire, Durham, New Hampshire, 03870, USA

* Corresponding author : francois.royer@unh.edu

1 **ABSTRACT**

2 We present in this paper a methodology for the non-linear analysis of long-term ecological
3 time series. Using a time-delay embedding procedure, we can investigate the complexity of 1-
4 dimensional signals and compare their structure to their stochastic counterparts. Recurrence-
5 based statistics and surrogate testing were used for this purpose. The method was firstly tested
6 on elementary models for illustrative purposes and its potentialities and limits reviewed. We
7 then investigated the time structure of 6 Mediterranean bluefin tuna trap catches that are
8 known to display long-term fluctuations. These time series showed significantly more
9 determinism than linear stochastic processes, displaying more recurring patterns/states than
10 stochastic surrogates. The short term structure of these series also showed a direct and
11 delayed negative feedback in their rate of increase. These results were homogeneous over all
12 the studied time series. An interaction between climate forcing and the life history traits of
13 bluefin tuna was postulated to explain the first result, while the density-dependent pattern may
14 be explained either by density dependence processes during early stages, or by non-successive
15 (non-yearly) spawning events in the Mediterranean.

16

17

18 **KEYWORDS**

19 *Thunnus thynnus*, recurrence, density-dependence, surrogate testing, embedding

1 INTRODUCTION

2 Studying variability in ecological systems over time and space is an extensive area of
3 research which has gained considerable interest in recent years (Gurney & Nisbet 1998,
4 Turchin 2003). Quantifying the relative importance of biotic and abiotic processes on free-
5 ranging marine populations has been, however, hampered by several limitations, among
6 which: (i) the lack of long ecological time series, (ii) the presence of observation errors,
7 especially when these populations are sampled blindly and (iii) the complexity of the
8 interactions between deterministic and stochastic mechanisms (Ripa et al. 1998, Cazelles &
9 Boudjema 2001, Fromentin et al. 2001, Turchin 2003). When conducting exploratory data
10 analysis, the alternation and combination of stochastic and deterministic periods, possibly
11 involving non-linear interactions and non-stationary noise, renders many standard linear
12 methods inadequate (such as the Fourier- and correlation-based approaches, Dale et al. 2002).
13 Moreover, fitting standard population models often supposes prior knowledge of the
14 underlying mechanisms and that any "uninformative" noise is simply additive (uncorrelated)
15 and can be assimilated to observation error. Recent studies have, however, demonstrated that
16 these assumptions are not often invalid: large-scale coherent patterns can emerge from
17 stochastic internal processes (e.g. from interactions between individuals or age classes,
18 Bjornstad et al. 2004), while environmental noise tends to be autocorrelated over space and
19 time (especially in marine ecosystems, Steele 1985, Vasseur & Yodzis 2004). Lastly, little
20 information may be available on the processes governing the population dynamics over the
21 studied time scales.

22 In this context, Atlantic bluefin tuna (*Thunnus thynnus*, L.) poses a challenging
23 problem. The exploitation of this pelagic migratory species dates back to the antiquity, and
24 has reached critically high levels during the past decades (see Fromentin & Powers 2005 for a
25 review). Despite continuous documentation efforts, detailed information on key processes (i.e.

1 migratory behaviour, population structure, habitat use, growth and reproduction) is still
2 lacking for elaborating and fitting an accurate population model. Furthermore, extreme
3 events, such as bursts and/or disappearance of bluefin tuna in given areas, have been reported
4 in the past (Tiews 1978, Mather et al. 1995), while apparent shifts in spatial distribution
5 and/or migration continue to be an actual and acute problem (Lutcavage et al. 1999,
6 Fromentin & Powers 2005). These observations suggest that the dynamics of bluefin tuna
7 may include several potential mechanisms operating at different spatial and temporal scales.
8 The recent collection of long time series from the coastal trap fisheries in the Mediterranean
9 greatly helped in describing patterns of variations, which were largely dominated by low-
10 frequency signals (Ravier & Fromentin 2001). These authors showed in particular that these
11 medium- to long-term fluctuations (~ 20 and $\sim 100-120$ years⁻¹) were synchronous amongst the
12 whole studied area (> 2000 km), while the inter-annual variability was only synchronous at a
13 local scale. Using linear regression techniques and stochastic simulation, Fromentin (2002)
14 and Ravier & Fromentin (2004) proposed a set of hypotheses to explain these medium- to
15 long-term fluctuations (from time-delayed effects related to a long life span and numerous age
16 classes to climate-induced effects), but they did not investigate the potential causes of inter-
17 annual variability.

18 In this paper, we perform a reanalysis of a set of these historical time series to: (1)
19 extract patterns of variability that may have remained hidden when using linear techniques
20 and (2) examine whether we could further discriminate between deterministic and stochastic
21 processes. We applied for this purpose a non-linear exploratory technique, namely Recurrence
22 Analysis (RA, Zbilut et al. 2002). RA is based on recent strategies derived from the signal
23 processing and chaos theory field and was first applied to various issues in optics, finance,
24 physiology, hydrology and chemistry (Eckmann et al. 1987, Marwan et al. 2002). It relies on
25 quantifying self-similarity in the data and detecting recurring patterns over time, with no

1 assumptions on their statistical nature. RA can thus characterize the dynamical properties of
2 the observed system, regardless of its non-linearity or non-stationarity (Gao & Cai 2000).
3 Such technique appears particularly suited for complex ecological systems for which little
4 mechanistic knowledge exist and where only one variable is measured (e.g. abundance or
5 catch, as it is the case here). There has been few applications of this technique in ecology:
6 Dippner (2002) employed RA to uncover deterministic (periodic) fluctuations and non-linear
7 interactions in a controlled mesocosm experiment, while Parrott (2004) showed how RA
8 could help in visually detecting disturbances in a simulated ecosystem.

9 As Recurrence Analysis is still relatively new to ecology, we here describe the basic
10 steps of this technique (i.e. phase-space embedding, recurrence plotting and recurrence
11 quantification) and then apply it to simple examples for illustration. We then show how the
12 use of surrogate data may help to extract ecological information from RA outputs. A specific
13 limitation of RA was identified in the case of finite, noisy data (a common situation in
14 ecology) and an additional step based on Partial Correlation (Berryman & Turchin 2001) is
15 proposed to address this issue. Finally, we apply RA on time series of bluefin tuna trap
16 catches and present results on previously hidden patterns. These new findings are discussed in
17 the light of current knowledge on bluefin's intriguing population dynamics and new modes of
18 variability are proposed.

1 MATERIALS AND METHODS

2

3 *Recurrence Analysis*

4 *Phase-space embedding*

5 Ecologists often study population dynamics through a small number of observed
6 variables, often no more than one (e.g. abundance, density or catch). Such data can be,
7 however, greatly informative under a basic assumption, i.e. that the sampled period is
8 sufficiently long to encompass most of the system's possible states. It is then possible to
9 reconstruct the dynamics of hidden processes from a single time series, thus uncovering some
10 aspects of their structure: this can be done using time-delay embedding or equivalently time-
11 delay coordinates. Let us write the population size at time t , N_t , as a function of its past
12 abundances (up to a delay d):

$$13 \quad N_{t+1} = f(N_t, N_{t-1}, N_{t-2}, \dots, N_{t-d}) \quad (1)$$

14 Such approach is common in quantitative ecology and can be seen as a generalized delay-
15 difference population model where neither the exact relationship f nor the order d are known.
16 If f is considered linear, an autoregressive model (of an estimated order d) could be fitted to
17 the data (Box & Jenkins 1976). In the non-linear case, embedding and recurrence analysis
18 provide a tool to characterise the complexity and shape of f (Boudjema & Chau 1996, Zbilut
19 et al. 2002).

20 We must first define what a “proper” embedding is. Given the time series of measures
21 N_t , with $t = 1, 2, 3 \dots T$, an embedding procedure will form a new series of points X_i of
22 coordinates $(N_i, N_{i+L}, \dots, N_{T-(m-1)L})$ with m the embedding dimension and L the lag. The
23 succession of X_i , now represents the multidimensional process of the time series as a
24 trajectory in a m -dimensional space, subsampled every L time step. The lag, L , or time delay,
25 corresponds to the memory of the process. L must be chosen so as to avoid neighboring points

1 being too close in the m -space and to maximize the information provided by each measure.
2 The embedding dimension m relates to the complexity (the order) of the hidden ecological
3 process. For example, a simple two-dimensional embedding will detect patterns in the growth
4 rate of a single population (Turchin 2003). Higher dimensions will reveal more complex
5 dynamics and interactions with other variables or pseudo-cyclic features (more points being
6 needed to “capture” such oscillations). Very high dimensions (15 to 20+) are indicative of
7 high order randomness, i.e. we would be unable to capture their dynamics through a finite
8 number of equations (see Hsieh et al. 2005). Takens' theorem (Takens 1981) states that
9 optimal m and L values will yield a one-to-one mapping between the reconstructed phase-
10 space and the phase-space of the hidden system. In other words, the dynamics of the
11 reconstructed system are the same as the dynamics of the original true system: characteristic
12 invariants are preserved (e.g. the number of controlling variables), thus allowing to identify
13 the possible processes underlying a noisy, apparently random data series. In practice, L is
14 chosen as the first minimum in the average Mutual Information function (which quantifies the
15 dependence between states taken at different lags), while m is derived using the Euclidean
16 False Nearest Neighbours method (see Kantz & Schreiber 1997).

17

18 *The Recurrence Plot*

19 A qualitative description of the observed dynamics can be obtained by plotting the (m -
20 dimensional) states that repeat themselves in the data and which are called recurrent states
21 (Eckmann et al. 1987). A Recurrence Plot (RP, Zbilut et al. 2002) is constructed by placing a
22 point at coordinates (i,j) on matrix of size $(T-(m-1)L, T-(m-1)L)$ whenever a state X_i on the
23 trajectory is considered close to another state X_j . The closeness between X_i and X_j is expressed
24 as the Euclidean distance $\|X_i - X_j\| < r$, where r is a fixed radius. The resulting matrix is
25 symmetrical (a point at (i,j) is also present at (j,i)), with the x and y axes being the new time

1 coordinates. RPs display patterns that can be qualitatively interpreted (Gao & Cai 2000,
2 Parrott 2004). For example, diagonals are formed by different parts of the trajectory evolving
3 similarly: such portions repeating themselves after a certain period of time indicate
4 deterministic behaviour. Horizontal and vertical lines indicate states that do not change over
5 successive time steps. Isolated dots, on the other hand, are likely to indicate states that repeat
6 themselves by chance (randomly). RPs remain, however, hard to read and a formal
7 description and quantification of these patterns is needed for further analysis. The value of r
8 may also affect their sensibility to observation noise, and the overall number of detected
9 recurrences: this could be a limitation, and we addressed it later in this section.

10

11 *Recurrence Quantification*

12 The structure of a Recurrence plot is usually defined through a number of statistics (Gao
13 & Cai 2000, Zbilut et al. 2002): 1) the Recurrence Rate (%REC) is the ratio of all recurrent
14 states to all possible states and is therefore the probability of the recurrence of a certain state
15 X_i ; 2) the Determinism (%DET) is the ratio of recurrence points forming diagonals to all
16 recurrence points, and shows the likelihood to have a piece of trajectory close to another one
17 at different times; 3) the Average Diagonal length (ADL) is the average time span they will be
18 close to each other and thus can be interpreted as a mean prediction time; 4) the Average and
19 Longest Diagonal Length (ADL and LDL): the inverse of the longest diagonal length is
20 related to the largest Lyapunov exponent of the system (a measure of its predictability,
21 Eckmann et al. 1987); 5) Laminarity (L) quantifies the amount of vertical and horizontal lines
22 in the RP, thus characterising the probability of having the system "locked" in a certain state;
23 6) The Trapping Time (TT) measures the average time spent in such laminar phases (Marwan
24 et al. 2002). These statistics (or probes, in the sense of Turchin 2003) are of great value for

1 characterizing the complexity of non-linear dynamics, as their interpretation is relatively
2 straightforward. Table 1 summarizes these 7 statistics and their interpretation.

3

4 *Hypothesis testing using surrogates*

5 The statistics describing the recurring points and patterns on an RP can then be used for
6 hypothesis testing against elementary random processes. The method of surrogates is used
7 here for the detection of determinism or highly recurring patterns. Linear surrogate techniques
8 have been introduced by Theiler et al. (1992) to determine whether experimental time series
9 are consistent with various classes of simple linear systems. Each surrogate is consistent with
10 a specific null hypothesis about the underlying system (e.g. a random variable or an
11 autoregressive-like process) while it retains in the same time some of the statistical features of
12 the original time series (e.g. mean and variance, power spectrum, complete empirical
13 distribution). Statistics are computed for the original data and the surrogates, thus allowing to
14 test (in a statistically rigorous way) if the data is consistent with the null hypothesis. This
15 method thus relies on both the algorithm employed to generate the surrogates and the
16 ecological relevance of the hypothesis one wants to test for characterising the system
17 (Royama 1992, Hilborn & Mangel 1997, Gurney & Nisbet 1998). The surrogates discussed in
18 the following section are based on various “colored” noise models (by analogy with the light
19 spectrum, see e.g. Halley 1996). “White noise” refers to a time series spectrum where all
20 frequencies are equally represented, meaning that successive values are uncorrelated. A “blue
21 noise” implies that the time series spectrum is dominated by short-term fluctuations (high
22 frequencies), whereas “reddened noise” are dominated by long-term fluctuations (low
23 frequencies). Some phenomena are known to have deeply reddened spectra, leading to 'brown'
24 or even 'black' noises (Cuddington & Yodzis 1999). Note that most environmental noises are
25 reddened, especially in marine systems (Steele 1985, Vasseur & Yodzis 2004).

1 The simplest null model would be that the data originates from “white noise” (i.e. an
2 independent and identically distributed variable, possibly non-gaussian). Corresponding
3 surrogates are often generated by simply shuffling the data, thus preserving the original
4 empirical distribution but destroying any time correlation. This algorithm is often referred to
5 as "type-0" in the literature (Theiler et al. 1992, Small & Tse 2003) and is mainly seen as an
6 autocorrelation test of the data prior to more advanced testing. It is worth noting that the white
7 noise null-hypothesis is generally unsuited to geophysical or ecological problems, since few
8 processes are able to generate white noise outputs (Hasselmann 1976, Rohani et al. 2004). In
9 ecological time series, white noise is often restricted to the observation process. However,
10 assuming a white noise for the dynamical process itself (e.g. population growth) is equivalent
11 to specifying the null-model of an almost perfectly compensated population (as the mean and
12 variance remain constant, see Royama 1992, Berryman & Turchin 2001).

13 A more complex null model ("type-1 surrogate") corresponds to the generation of
14 “reddened noise”, which is the usual way of trying to capture endogenous fluctuations in
15 populations (Royama 1992). In such case, the linear properties of the time series are fully
16 described by its autocorrelation function or equivalently by its power spectrum. Surrogate
17 data are therefore build by adding random phases in $[0,2\pi]$ to the components of the Fourier
18 transform of the observed time-series, and then computing its inverse Fourier transform
19 (Theiler et al. 1992). The resulting surrogates are Gaussian and have the same mean, variance
20 and power spectrum as the original data. Applying this algorithm to a time series of
21 abundance is particularly interesting to detect regulation (i.e. density dependence) in
22 population dynamics, as it is equivalent to the null model of an unregulated population growth
23 (Berryman & Turchin 2001). Testing against coloured noise further allows to assess the
24 significance of pseudo-oscillations/trends in the data, since univariate autoregressive
25 processes cannot *per se* support deterministic low-frequency oscillations (Rohani et al. 2004).

1 Statistical testing of series with reddened spectrum is still the subject of research and debate
2 in ecology, as many tests rely on the assumption of white noise and lose their power for
3 autocorrelated or non-stationary series (see Morales 1999, Rudnick & Davis 2003).

4 An even more complex null model ("type-2 surrogate") will preserve both the original
5 non-Gaussian distribution and the power spectrum of the data (Venema et al., unpub. data).
6 Such surrogates can be generated using the Amplitude Adjusted Fourier Transform (AAFT)
7 algorithm. Such calculation slightly distorts the spectrum, but generally yields to satisfying
8 surrogates for sufficiently long data sets, e.g. more than 100 points (Schreiber & Schmitz
9 1996). This is equivalent to producing non-Gaussian data from an autoregressive process
10 combined with a monotonic and static non-linear observation function. Strictly speaking, any
11 non-linearity here is non-dynamic, i.e. attributed to the measurement function. Such
12 formulation is very convenient in time-series analysis, and ecological studies in general,
13 where a major task is to separate sampling error from dynamical noise (Turchin 2003).

14 For each test, the generation of a large number of surrogates allowed to derive a
15 distribution of the RP-based statistics. As most distributions were non-Gaussian, the
16 significance level was assessed in a non-parametric manner, i.e. by counting up the fraction of
17 realizations that produce a value greater than the observed one. The threshold was chosen
18 greater or equal than $2\alpha^{-1}-1$, where α is the residual probability of false rejection (see Marwan
19 2003). All the calculations were performed using Matlab 6.5 (Matlab 2003) and the
20 calculation of the RPs statistics were achieved using code made available by Marwan et al.
21 (<http://www.agnld.uni-potsdam.de/~marwan/toolbox/>) and Venema (<http://www.meteo.uni-bonn.de/mitarbeiter/venema/themes/surrogates/>).

22

23
24 *Observation noise and the choice of r: limitations and solutions*

1 We noted that time-delay embedding allows to retrieve the real dynamics of a partially
2 observed system in the noise-free case or for true (additive) observational noise only, but
3 there is no guarantee that noise-driven dynamics will be completely reconstructed. Previous
4 works have, however, shown that this technique can be powerful for diagnostic and
5 identification purposes, when there is only little knowledge about the underlying processes of
6 a system (Boudjema & Cazelles 2001).

7 In the application case of finite, noisy time series (a common feature in marine ecology
8 where observation errors are common), the choice of r may limit the detection of high
9 frequency patterns: for example, two recurring system states X_i and X_j could be infinitely
10 close, but measurement error combined with a small r value would misclassify them as non-
11 recurring. On the other hand, a more permissive (larger) r value would increase the number of
12 apparent recurring states in the system, and thus the risk of false detections. We addressed this
13 by: (i) using the same (arbitrary small) r values for all series and corresponding surrogates,
14 and (ii) completing our analysis with the Partial Rate Correlation Function (PRCF, Berryman
15 & Turchin 2001), analogous to a 2-dimensional embedding in the RA framework. The PRCF
16 has been elaborated to detect density-dependent structures (thus high-frequency processes),
17 by investigating the (d -lagged) correlations between the rate of change $R_t=N_t-N_{t-1}$ and R_{t-d}
18 with the effect of earlier lags removed. Computing the autocorrelation of the rate of change
19 (R_t) rather than of the abundance (N_t), as with the Partial Autocorrelation Function (PACF),
20 avoids any masking effects because of positive autocorrelation due to reproduction (Berryman
21 & Turchin 2001). Its significance can be assessed directly without bootstrapping, using
22 Bartlett's criterion, $\pm 2n^{-1/2}$, with n the length of the series, (Box & Jenkins 1976, Royama
23 1992). As seen above, the null model of the PRCF is a population subject to unbounded
24 growth (i.e. a reddened noise), which should be preferred to the PACF null model (i.e. white
25 noise). This would be further consistent with the bluefin historical catches, which display

1 long-term (reddened) fluctuations (Ravier & Fromentin 2001). This, and its ability to uncover
2 short-scale patterns (typically over a few lags) made us use the PRCF as a complimentary step
3 in the RA analysis of these series.

4

5 **Data**

6 *Simulated examples: three elementary stochastic models*

7 For illustration and testing purposes of the Recurrence Analysis, we firstly simulated
8 time series from 3 elementary stochastic and non-linear models, that are currently used or
9 referred to in ecology. The first model is a simple autoregressive process generating a random
10 walk (i.e. a brown noise):

$$11 \quad N_{t+1} = N_t + \varepsilon_t, \quad (2)$$

12 where ε_t is a gaussian white noise. This model simply corresponds to a population subject to
13 unbounded growth (Fig. 1a) and therefore defines the null model in the sense of Berryman &
14 Turchin (2001). The second time series was derived from a stochastic order-2 autoregressive
15 model:

$$16 \quad N_{t+1} = \exp(aN_t + bN_{t-1} + \varepsilon_t) \quad (3)$$

17 This model depicts how a delayed autoregressive process can be affected by an observation
18 process of the exponential family (Fig. 1b). Non-linearity is attributed here to the exponential
19 function which amplifies the random fluctuations of the autoregressive process, and would
20 require log-transforming the series based on some *a priori* assumptions (Fromentin et al.
21 2001). Lastly, a third non-linear time series was constructed by adding gaussian noise to a
22 simple two-species trophic model, classically called NPZ model in plankton ecology (e.g.
23 Franks et al. 1986):

$$24 \quad N = N_{tot} - P - Z \quad (4)$$

$$1 \quad \frac{dP}{dt} = \frac{uPN}{N + N_0} - g \frac{ZP}{P + P_0} \quad (5)$$

$$2 \quad \frac{dZ}{dt} = ag \frac{ZP}{P + P_0} - dZ \quad (6)$$

3 Where N is the nutrient concentration, and P and Z the phytoplankton and zooplankton
4 abundance, respectively. Parameters are set as follows: $N_{tot} = 5$ (initial nutrient concentration),
5 $N_0 = 0.1$ (nutrient threshold), $P_0 = 0$ (initial phytoplankton abundance), $u = 0.3$ (uptake factor
6 or growth rate of P), $g = 0.2$ (grazing factor), $a = 0.4$ (grazing conversion factor), $d = 0.08$
7 (death rate of Z). The model is supposed to be partially observable, *i.e.* only the
8 phytoplankton variable P is measured, and blurred with white observation noise $w \sim N(0,1)$.
9 This parameterization generates an oscillatory behaviour in the system and a series of cycles
10 in the measured (P) time-series (Fig. 1c).

11 The embedding procedure implies that the full complexity of the system may be
12 retrieved from the data sets. In other words, time series must be long enough for the results to
13 be robust. We therefore tested RA with various lengths of each simulated time series, from 50
14 points up to 500 points.

15

16 *Historical time series of bluefin tuna catches*

17 Time series of bluefin tuna catches (in number of fish) have recently been collected
18 after an intensive work on archives and historical material (Ravier & Fromentin 2001). These
19 were derived from the Mediterranean and Atlantic costal trap fishery, which used a fixed gear
20 to intercept bluefin tuna during their spawning migration (Fromentin & Powers 2005). This
21 original dataset spans a 360-years period, from 1600 to 1960, with notable gaps between 1820
22 and 1850. The configuration and mode of operation of these traps remained quite stable
23 during that period (Farrugio 1981), and they all displayed synchronous variations at dominant
24 periods of ~ 20 and ~ 100 -120 years, the latter accounting for 50-70% of the total variance.

1 This indicated that a common process was likely to be observed, and that trap catches are
2 likely to reveal long-term fluctuations in the abundance of migrating bluefin tuna (Ravier &
3 Fromentin 2001, 2004). Although a synthetic, full-length, time series was computed by these
4 authors, it was rather chosen to apply RA to the original (shorter) series, as recurring patterns
5 are likely to have been smoothed or lost during this process. Following the tests on the
6 minimum needed length and Ravier & Fromentin (2004), we selected time series displaying at
7 least 125 contiguous years without missing values (Fig. 2). The 6 retained series came from
8 the Sicilian traps (Formica, Favignana and Bonagia: 180 annual values of catches from 1634
9 to 1813) and from the Sardinian traps (Porto-Paglia, Porto-Scuso and Isola-Piana: 136 values
10 from 1825 to 1960). The unusual length of these time series also allowed us to assume the
11 bluefin population visited a large range of abundance levels and patterns (i.e. of states), which
12 is a prerequisite for phase-space reconstruction and non-linear analysis.

1 RESULTS

2 *Simulated examples: analysis of elementary stochastic models*

3 For comparison purposes, the three test time series were embedded in phase space of
4 dimension 2 and unit time delay. The corresponding recurrence plots (Fig.1) are easily
5 distinguishable, displaying a number of vertical, horizontal and diagonal features. The random
6 walk model is characterized by checker-board like patterns (Fig. 1d), while the exponential
7 autoregressive process tends to fill up the RP with neighbouring points (Fig.1e). While the
8 deterministic fluctuations of the P compartment of the last model are clearly visible in the
9 form of sharp structures, once blurred by observation noise, they become more similar to the
10 squared patterns of the two previous models (Fig.1f). Still, the succession of the three cycles
11 remains visible, separated by horizontal and vertical white strips indicating abrupt variations
12 in the state vector.

13 We then tested these 3 elementary time series against the different types of surrogates
14 using an hierarchical approach, i.e. we firstly tested against “type-0” surrogates, if H_0 was
15 rejected we then tested against “type-1” and if it was rejected again we tested against “type-2”
16 surrogates (all the results being summarised in Table 2). Note that a relatively low number of
17 surrogates (20 to 30) proved to be sufficient for getting reliable results.

18 The test of the random walk model, that mostly displays long-term fluctuations,
19 against “type-0” surrogates (shuffling the data in time domain) led to reject the null model in
20 all the cases. When these surrogates were then generated by randomly shuffling the phases of
21 the Fourier transformed data (“type-1” surrogates), the RP-based statistics were all
22 indistinguishable from their surrogate counterparts, except for the Laminarity index (using the
23 non-parametric test introduced earlier, $2\alpha^{-1}-1$ with $\alpha = 0.05$, see Table 2). “Type-1” surrogates
24 allow us to correctly identify the stochastic nature of the random walk and, thus, to avoid to

1 attribute these pseudo-cycles to deterministic processes or external forcing (as it could be
2 tempted at first sight).

3 The exponential of an autoregressive process, that appears close to white noise but
4 further exhibits occasional bursts (Fig. 1b, 1e), led to reject the null model based on “type-0”
5 or “type-1” surrogates with high confidence ($p < 0.05$ for %REC, %DET, LDL, ADL, L and
6 TT). This indicates that such a time series could not be assimilated to a “white” or a
7 “reddened” noise. However, the test against “type-2” surrogates was not rejected, each
8 statistic being indiffereniable from their surrogate counterparts. So the exponential of an
9 autoregressive process was consistent with red noise (unregulated growth) combined with a
10 non-linear observation function (i.e. a static non-linearity).

11 The NPZ model is a mix of deterministic and stochastic processes (Fig. 1c). “Type-0”
12 surrogates (time shuffling) retain the time distribution but destroy both the frequency
13 distribution and any dynamic structure (Fig. 3a). The RP of such surrogates also appears quite
14 different from this of the NPZ model (Figs. 1f, 3d) and as expected this null model was
15 rejected for all the statistics (Table 2). “Type-1” surrogates (phases shuffling) retain the
16 frequency distribution (Fig. 3b) and display pseudo-cycles similar to those observed in the
17 NPZ model. However, such variations do not appear as regular as in the original data and the
18 RPs appear also different (Figs. 1f, 3e). “Type-2” surrogates retain the frequency distribution
19 and the time distribution (Fig. 3c). Nonetheless, its corresponding RP also appeared different
20 from original data and both null models (“type-1” and “type-2”) were rejected for the NPZ
21 model for some statistics (Fig. 3f). Significant more recurring patterns (%REC) than red noise
22 surrogates, with a higher prediction time (%LDL) were indeed found (Table 2). This indicates
23 that the succession of the three cycles (visible on the RP, Fig. 1f) did not result from
24 stochastic processes but from deterministic ones.

25

1 *Recurrence analysis of the bluefin tuna time series*

2 Given the global synchrony between the Mediterranean traps (Ravier & Fromentin
3 2001), embedding parameters were set identical for the 6 time series. False nearest neighbours
4 advocated for an embedding dimension m of about 8, while the first minimum in the self-
5 mutual information function favoured small time-delays (i.e. $L \sim 1$ to 3). As the fishing
6 activity occurred every year during spawning migration (i.e. May to July), it can be
7 hypothesized that the frequencies of the observation and the ecological processes are closely
8 matching: we therefore used a time delay L of 1. This avoided sub-sampling the data (the full
9 time series were used), but the risk is to include observation noise in the embedding process.
10 Unsurprisingly, the null hypotheses of white noise (i.e. “type-0” surrogates) was rejected for
11 all time series and for all recurrence statistics. Less trivially, the same results were obtained
12 with reddened noise (i.e. “type-1” surrogates). In other words, the catch time series showed
13 significantly more recurring patterns than implied by a white null model (uncorrelated random
14 fluctuations) or a reddened null model (autoregressive model). To illustrate this result,
15 examples of surrogates are given for the Formica time series in Fig. 4. “Type-0” and “type-1”
16 surrogates appear indeed quite different of the original structure of the observed time series,
17 but “type-2” surrogates look more similar. However, the null model that preserves both the
18 original non-gaussian distribution of the trap time series and their reddened power spectra
19 (“type-2” surrogates) only allowed to uncover a part of recurring features of the various trap
20 time series, but not all (Table 3). All the time series indeed displayed significant differences
21 ($p < 0.05$) with the null model for given statistics (being, however, different among the traps).
22 For example, Formica, Favignana, Bonagia and Piana displayed significantly more recurring
23 vectors than the null model (Table 3). Favignana and Piana also displayed more deterministic
24 structures and Favignana and Bonagia larger laminar time periods than what would be
25 expected under the null model. Note also that Bonagia further exhibited a greater average

1 diagonal length, that could indicate a longer mean prediction time. Porto-Scuso further
2 displayed a significantly larger prediction horizon (LDL), maybe indicating dynamics of
3 lower complexity. Although the 6 trap time series displayed different results, Favignana and
4 Bonagia displayed quite similar statistics, being different of the third Sicilian trap (Formica)
5 and the three Sardinian ones. Among the Sardinian traps, Porto-Paglia and Porto-Scuso
6 appear the closest ones, while Isola Piana displayed some recurrent patterns close to the
7 Sicilian traps.

8

9 *Partial Rate Autocorrelation Function*

10 To complete the recurrence analysis, the Partial Rate Autocorrelation was computed for
11 the first-order finite difference of all six series. The PRCF plots displayed negatively
12 correlated autocorrelation in R_t over two successive time lags (i.e. of order 2). This result was
13 true for the six catch time series considered here (Fig. 5), as well as the other smaller series in
14 the original data set (results not shown). The fact that significant negative autocorrelation is a
15 common feature of all trap catch time series would, thus, strongly advocate for density-
16 dependent structures in migrating bluefin tuna. The occurrence of density dependence could
17 explain why we previously found more recurring/deterministic structures than the “type2”
18 null model in several time series (see above or Table 2). Weakly significant correlations could
19 be detected at other lags according to Bartlett’s criterion (e.g. lag 11 for Formica, 10 for
20 Bonagia, 4 for Porto Paglia, 3 for Porto Scuso, and 3, 7 and 10 for Isola Piana). However,
21 only the negative direct and delayed correlations (lags 1 and 2) appeared significant for all
22 time series.

1 **DISCUSSION**

2 *Usefulness and significance of recurrence analysis in ecology*

3 The recurrence statistics proposed by Zbilut (2002) and Marwan (2002) can be seen as
4 new "probes" of complexity (sensu Turchin 2003) for ecological time series. Recurrences can
5 be used on raw data analysis and are complementary to standard time series tools, such as
6 autocorrelation function, Fourier spectrum, which provide clues about the ecological
7 "smoothness" of underlying processes (Chatfield 1999). According to the algorithmic
8 definition of randomness (Chaitin 2001), recurring patterns are highly informative about the
9 degree of stochasticity in a data set. However, the qualitative information produced by the
10 Recurrence Plot, although useful, remains insufficient (and somewhat too subjective) to
11 interpret time series. Therefore, recurrent analyses must also include the computation of
12 RP-based statistics. We further think that these statistics have to be tested against clear null
13 models and we, therefore, proposed three types of surrogates. The use of surrogates indeed
14 provide a convenient answer to the question: "how much deterministic structures is embedded
15 in a time series?". In that sense, surrogate-based RA is nothing more than hypothesis testing,
16 i.e. we assume a "trivial" option to be true unless we can reject it by finding significant
17 structure in the data. Using three elementary and simulated time series, we showed that this
18 procedure was quite powerful and able to distinguish between pure stochastic process (i.e. the
19 random walk) and the mixture of stochastic and deterministic ones (i.e. the noisy
20 autoregressive model and the NPZ model). Doing so, we can then identify the most likely
21 models to an ecological time series. For instance, a state-space formulation may be relevant
22 for the noisy autoregressive example, which may not be obvious at first glance (i.e. the bursts
23 could be indeed interpreted as dynamic non-linearity). Finally, we are limited by the degree of
24 complexity of the null hypotheses that we can propose. To date, the simplest (or most trivial)
25 processes we are able to generate by constrained simulation belong to the linear or weakly

1 non-linear class (e.g. surrogates of “type-0, 1, 2”). Recent efforts on constrained
2 randomization broaden the repertoire of such null hypotheses. Generally, random processes
3 blur deterministic dynamics through observation noise (generally additive), and dynamical
4 noise (possibly propagated through the system). Quantifying the resulting variability through
5 RA can help in separating both sources, acknowledging that no technique is noise-insensitive.
6 RA is subject to such noise through the choice of r : for our specific application on bluefin
7 tuna, we used the Partial Rate Autocorrelation Function to complete the analysis. Knowing
8 that ecological time series are often of limited length (especially with *in-situ* data), we tested
9 the sensitiveness of the RA analysis to time span. Using simulated data, we, thus, repeated the
10 RA steps after truncating a time series of 500 values in sub-samples of length 50, 100, 150,
11 200, 300 and 400. Boxplots of the results (one plot for each RP-based statistic) were similar
12 until a length of ~100-150 values (after what some structures were lost). However, such a
13 limit is just indicative and depends on the ratio of deterministic structures *vs* noise. Most RP
14 statistics appear to be invariant in time or displayed a plateau-like behavior, except for the
15 statistics associated to the length of some event (e.g. the LDL-Longest Diagonal Line and the
16 Laminarity), which grew almost linearly with the series length, especially for the random
17 walk model (results not shown).

18

19 *Historical fluctuations in bluefin catches*

20 Recurrence Analysis was applied to the historical time series of bluefin catches from the
21 Mediterranean coastal traps. A dimension of 8 was found for phase space embedding, while
22 testing against surrogates indicated more recurring features than expected from a linear
23 stochastic process. While the lengths of these series were at the lower limit of what is needed
24 for reliable results, the results among the six chosen series were globally homogeneous and
25 significant. This analysis, thus, confirms, through a quantitative and objective test, the non-

1 stochastic nature of the long-term fluctuations in bluefin tuna trap catches, as Ravier &
2 Fromentin (2001, 2004) postulated it intuitively. In particular, hypothesis testing using
3 Fourier-based surrogates revealed that the observed dynamics could not be described solely
4 by their frequency spectrum and time distribution: most time series displayed more recurring
5 patterns than expected under this null hypothesis.

6 These results are further in agreement with the general view that marine populations
7 display non-linear stochastic variations (Sugihara 1994, Hsieh et al. 2005). In particular,
8 Hsieh et al. (2005) showed that time series of commercial landings of Pacific salmon and
9 trout and of zooplankton (obtained from CalCOFI surveys) were characterized by embedding
10 dimension of 3 to 8, an indication of low dimensional non-linear processes. On the other
11 hand, physical variables such as sea surface temperature and climatic oscillation indices were
12 characterized by high and very high dimensions (10 to 20), indicating that high-order linear
13 stochastic processes are better at capturing their variations. A dimension of 8 for the bluefin
14 tuna time series is coherent with the range computed by these authors for their biological
15 series and would advocate for low dimensional non-linear processes. This view is in partial
16 agreement with the hypothesis of “cohort resonance”, as a possible mechanisms to generate
17 low frequency oscillations when many age classes interact (Fromentin 2002, Bjornstad et al.
18 2004). Following this, bluefin tuna population would act as low-pass filter (a smoother) of
19 variability affecting critical processes, such as recruitment or growth at early stages. However,
20 such autoregressive process still posses a linear basis and hardly supports deterministic
21 (recurrent) patterns and rather produces random long-term variations. This hypothesis can be
22 extended in the light of our results: it is suggested that a blend of environmental forcing and
23 non-linear biological response occurs at time scales of 10-30 years, which is of the order of
24 the life span of Atlantic bluefin tuna (Mather et al. 1995). Atmospheric forcing (synthesized
25 through e.g. the NAO) could therefore affect bluefin population through the occurrence and

1 strength of meso-scale processes where spawning is known to occur (Garcia et al. 2003). As
2 bluefin tuna displays a “homing-like” behavior and quickly switches from a wide
3 geographical distribution to restricted spawning areas (Fromentin & Powers 2005), such
4 atmospheric forcing is likely to induce non-linear responses. These delayed interactions may
5 remain hidden to classic correlation-based analysis and would therefore need to be
6 investigated through mechanistic age-structured models. Possible interactions with local units
7 in a meta-populations context could also lead to non-linearity in the response to climate
8 forcing (Fromentin & Powers 2005).

9 As Recurrence Analysis may yield less reliable results at very small scale (due e.g. to
10 high observation noise), we completed the methodology with the computation of Partial Rate
11 Autocorrelation Functions (a simplified form of phase-space embedding). A significant
12 negative feedback was uncovered for all six time series, at lags 1 and 2. Such feedback is
13 usually defined as a density-dependence in ecological theory and seen as a necessary
14 condition for population regulation in general (Berryman 2002). Density dependence theory
15 supposes a negative relationship between population abundance (N_t) and its growth rate (R_t , a
16 function of per-capita birth, death, immigration and emigration rate). As Atlantic bluefin tuna
17 spawn in restricted areas in the Mediterranean (Fromentin & Powers 2005), one could first
18 see resource/space limitations as a possible explanation (bluefin larvae reaching an active and
19 voracious swimming stage after only 2 weeks, see Mather et al. 1995) or even through
20 cannibalism (within early stages or from adults on age-0 classes). High densities of Atlantic
21 bluefin tuna larvae or cannibalism have not been reported, but such observation is difficult to
22 collect and remains scarce (Garcia et al. 2003, Fromentin & Powers 2005). Note, however,
23 that high density of tuna larvae has been already observed for a neighbouring species, the
24 southern bluefin tuna (T. Smith Pers. Comm.). Density-dependent processes appear to be less
25 probable within the adult stages because of their wide spatial distribution and highly

1 migratory behavior, so that local depletion in forage or changes in predation (e.g. by killer-
2 whales) are more likely to induce changes in migration patterns than in abundance (e.g. Tiews
3 1978). The naturally high unpredictability of small pelagics populations (Spencer & Collie
4 1997) also hardly supports such deterministic hypothesis. An alternative to density-
5 dependence in early stages would involve non-yearly spawning (a feature that has been
6 observed on bluefin tuna in captivity, see Lioka et al. 2000). According to the observed order-
7 2 negative feedback, a significant portion of migrants coming from the Atlantic would thus
8 spawn in the Mediterranean every three years only, which can be caused either by: (1) a
9 physiological inability of adults to perform gonad maturation every year or (2) the
10 frequentation of other spawning grounds (as already postulated by various authors, e.g.
11 Lutcavage et al. 1999, Ravier & Fromentin 2004). This hypothesis can be further interpreted
12 in the framework of marine metapopulations, a concept shown to be highly relevant in the
13 case of bluefin tuna (Fromentin & Powers 2005). Synchrony within the fraction of migrants
14 may then be explained through stabilizing mechanisms similar to the “school-trap hypothesis”
15 (Bakun & Cury 1999). There is however little knowledge on the maturation costs of gonads in
16 spawning bluefin. While there is evidence of rapid maturation in relation to warmer
17 temperatures once in the Mediterranean (Medina et al. 2002), year-to-year changes have not
18 been reported and may lack physiological basis (Lutcavage, pers. comm.). Such delayed
19 negative feed-back remains nonetheless highly significant, and hypotheses need now to be
20 built and tested using alternative sources of data.

1 **CONCLUSION**

2 Disentangling the effects of endogenous and exogenous processes in ecology is an
3 important field of research. We presented in this work a method that can uncover
4 deterministic features in ecological time series. Recurrent features are part of the most basic
5 clues about the structure of a biological system (e.g. reproduction patterns). Using phase-
6 space embedding and recurrences as a signature for determinism, we could test whether
7 observed fluctuations are consistent with linear stochastic surrogates. Such analysis can
8 provide clues on the underlying complexity of measured signals (e.g. whether a linear or non-
9 linear model should be used, whether internal stochastic processes suffice to induce the
10 observed variability...). The application of this method to historical time series of bluefin
11 catches revealed that the long-term fluctuations could not be assimilated to simple stochastic
12 or autoregressive processes. Small scale variability was addressed using Partial Rate
13 Autocorrelation Functions and we further uncovered negative feedback of order 1 and 2 in all
14 time series, indicating either density dependence processes during early stages, or non-yearly
15 spawning. We advised for further work in the modelling of these patterns. In particular, the
16 mechanistic modelling of bluefin tuna response to oceanic variability should be considered,
17 using age structured and time delay formulations. The origin of the density-dependent pattern,
18 on the other hand, should be investigated in a hypothesis testing approach.

1 **ACKNOWLEDGEMENTS**

2 The authors are thankful to four anonymous referees that helped in improving the clarity of
3 this manuscript. Bernard Cazelles also provided valuable and helpful comments to an earlier
4 version of the ms. This work was conducted as part of FR PhD project, co-funded by CLS
5 (Collecte Localisation Satellites, Ramonville-St-Agne, France) and IFREMER (Institut
6 Francais de Recherche pour l'Exploitation de la Mer).

7

1 **TABLES AND FIGURES**

2

3 Table 1. Name, acronym and meaning of the statistics used in the Recurrence Analysis (from
4 Gao & Cai 2000, Zbilut et al. 2002).

5

6 Table 2. Significance (at a 95% level) of the recurrence statistics for three elementary models
7 (see text). Each test was performed using an embedding dimension $m = 2$ and a lag $L = 1$.

8

9 Table 3. Significance (at a 95% level) of the recurrence statistics for the six bluefin catch time
10 series from the coastal Mediterranean traps. Each test was performed against surrogates of
11 type-2 (i.e. a stochastic autoregressive process with static non-linearity), using an embedding
12 dimension $m = 8$ and a lag $L = 1$.

13

14 Figure 1. Time series of the three elementary stochastic models: a) Random Walk, b)
15 exponentially transformed autoregressive process, c) noisy NPZ model, together with their
16 corresponding recurrence plot (d,e,f) computed with an embedding dimension of 2 and time
17 delay 1. On the recurrence plot of noisy NPZ model (1f) is further inserted the recurrence plot
18 of non-noisy (i.e. fully deterministic) NPZ model.

19

20 Figure 2. Historical time series of bluefin tuna fluctuations as derived from the Sicilian and
21 Sardinian traps (after Ravier & Fromentin 2001). Arrows show the location of each group of
22 traps.

23

24 Figure 3. a) “type-0” (time shuffled), b) “type-1” (phase-shuffled), c) “type-2” (constrained
25 phase randomised Fourier transform) surrogates corresponding to the noisy NPZ model.

1 Corresponding recurrence plots (d-f) were computed with an embedding dimension of 2 and
2 time delay 1.

3

4 Figure 4. a) Recurrence plot of the Formica time series with an embedding dimension of 8
5 and a time delay of 1. Examples of surrogates are given on the right showing, b) the original
6 log-transformed data time series, c) a corresponding “type-0” surrogate (time shuffled), d)
7 corresponding “type-1” surrogate (phase-shuffled) and e) corresponding “type-2” surrogate
8 (constrained phase-randomised Fourier transform).

9

10 Figure 5. Partial Rate Correlation Functions (PRCFs) for the 3 Sicilian and 3 Sardinian time
11 series: Formica, Favignana, Bonagia, Porto-Paglia, Porto-Scuso and Isola-Piana.. The
12 confidence band (grey area) is given by Bartlett's criterion ($\pm 2n^{-1/2}$).

13

1 **Tables**

Name	Acronym	Interpretation
Recurrence Rate	%REC	Ratio of all recurrent states to all possible states
Determinism	%DET	Ratio of diagonal recurrent states to all recurrent states
Average Diagonal length	ADL	Mean prediction time of the time series
Longest Diagonal length	LDL	Largest prediction time of the time series
Laminarity	LAM	Amount of laminar phases (same state occurring consecutively)
Trapping time	TT	Average time spent in laminar phases

2

3

Table 1

4

	Model 1 (RW)			Model 2 (expRW)			Model 3 (NPZ)		
	Type-0	Type-1	Type-2	Type-0	Type-1	Type-2	Type-0	Type-1	Type-2
%REC	*	-	-	*	*	-	*	*	*
%DET	*	-	-	*	*	-	*	*	*
%ADL	*	-	-	*	*	-	*	*	-
LDL	*	-	-	*	*	-	*	*	*
LAM	*	*	-	*	*	-	*	-	-
TT	*	-	-	*	*	-	*	-	-

5

6

Table 2

7

Statistic	Formica	Favignana	Bonagia	Paglia	Scuso	Piana
%REC	<i>0.04**</i>	<i>0.00**</i>	<i>0.00**</i>	<i>0.16</i>	<i>0.14</i>	<i>0.02**</i>
%DET	<i>0.11</i>	<i>0.04**</i>	<i>0.16</i>	<i>0.47</i>	<i>0.19</i>	<i>0.04**</i>
ADL	<i>0.47</i>	<i>0.11</i>	<i>0.04**</i>	<i>0.28</i>	<i>0.35</i>	<i>0.28</i>
LDL	<i>0.39</i>	<i>0.11</i>	<i>0.18</i>	<i>0.12</i>	<i>0.02**</i>	<i>0.76</i>
LAM	<i>0.72</i>	<i>0.00**</i>	<i>0.00**</i>	<i>0.11</i>	<i>0.25</i>	<i>0.28</i>
TT	<i>0.61</i>	<i>0.06</i>	<i>0.14</i>	<i>0.11</i>	<i>0.09</i>	<i>0.67</i>

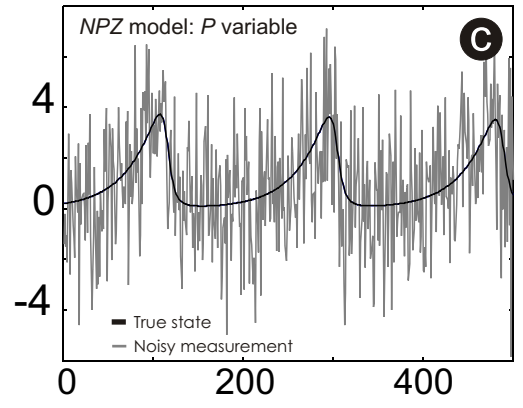
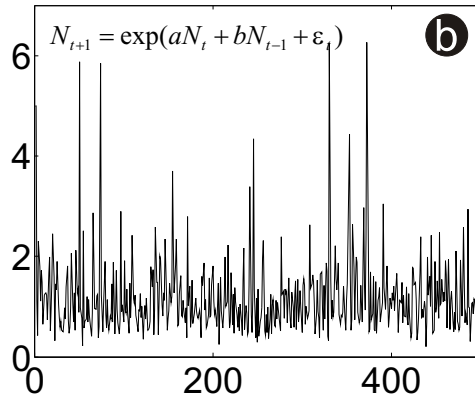
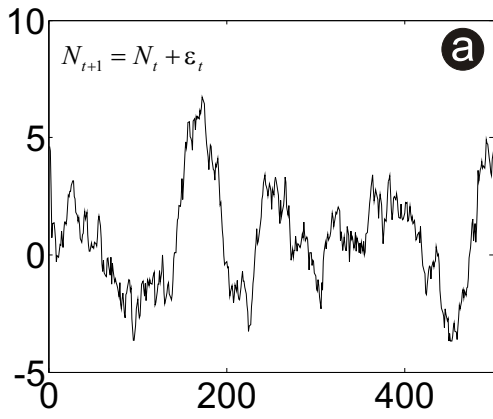
8

9

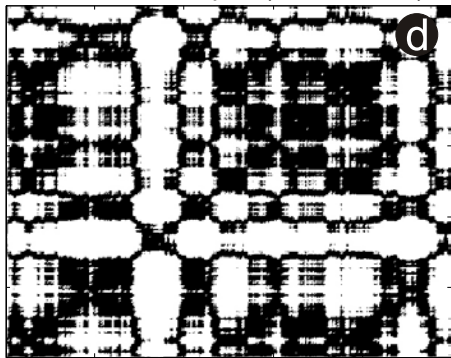
Table 3

10

11



Recurrence plot ($m = 2, L = 1$)



Recurrence plot ($m = 2, L = 1$)

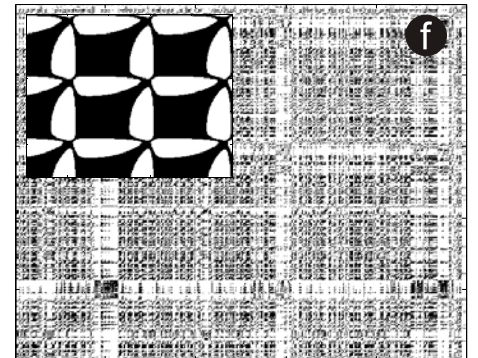
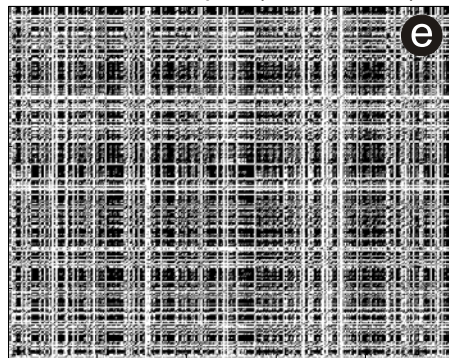


Figure 1

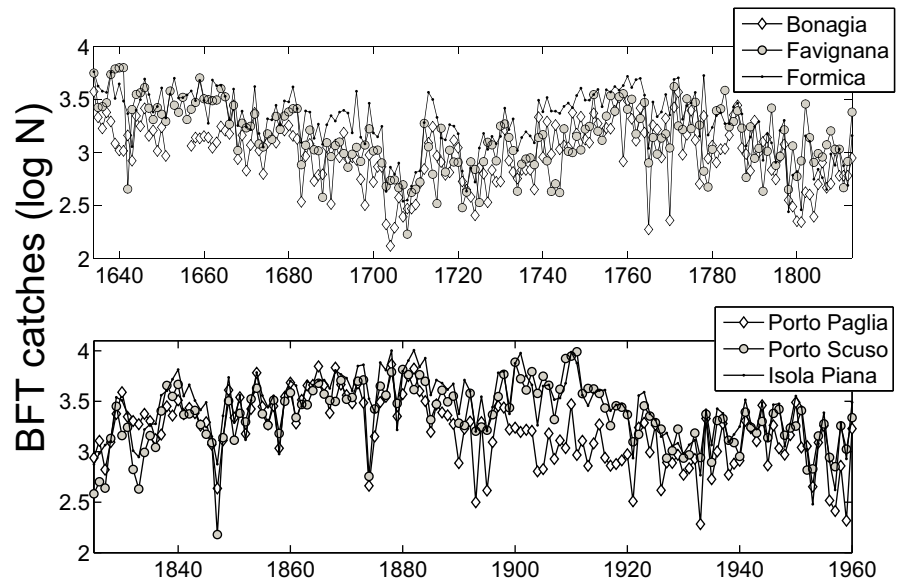
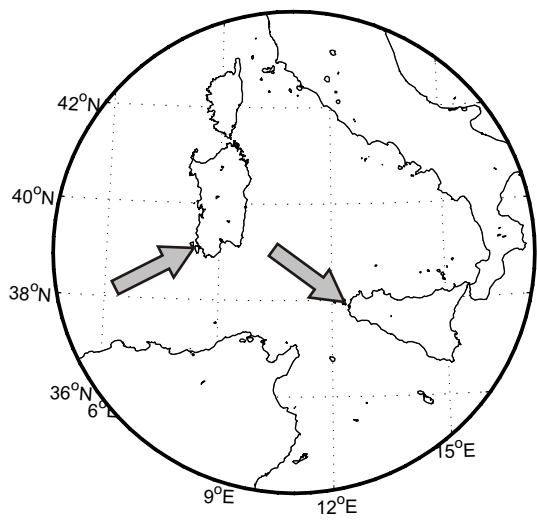
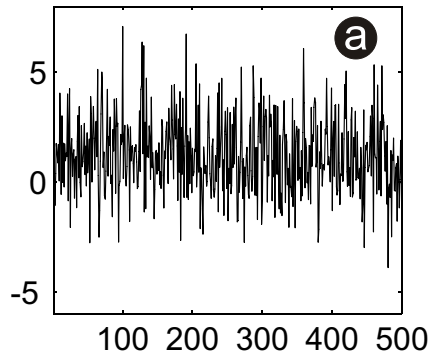


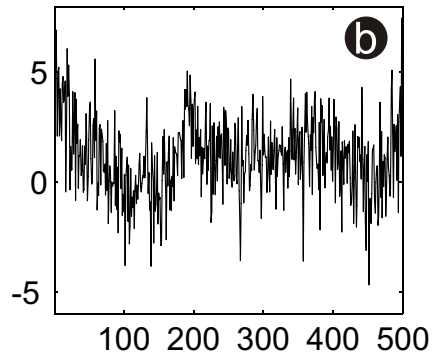
Figure 2

NPZ model

Type-0 surrogate



Type-1 surrogate



Type-2 surrogate

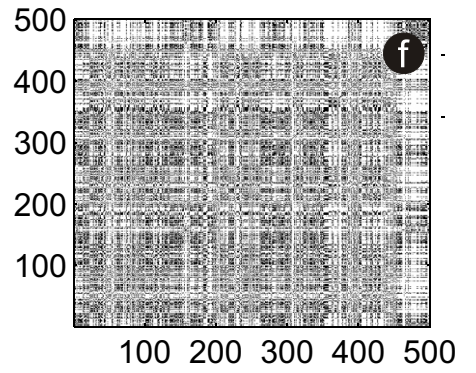
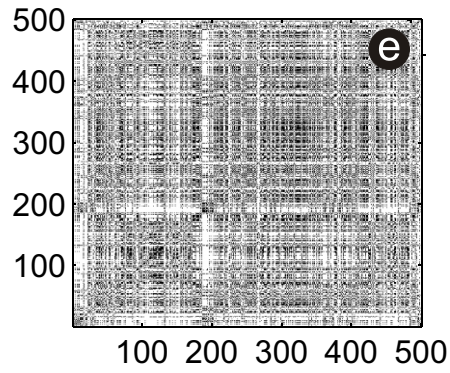
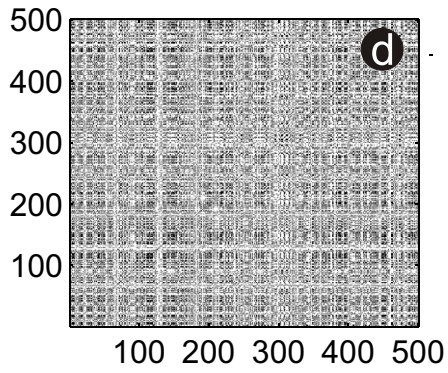
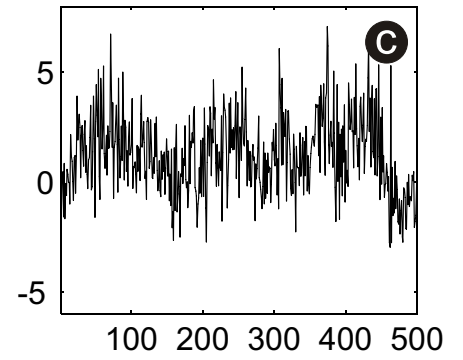


Figure 3

Recurrence plot: Formica ($m = 8, L = 1$)

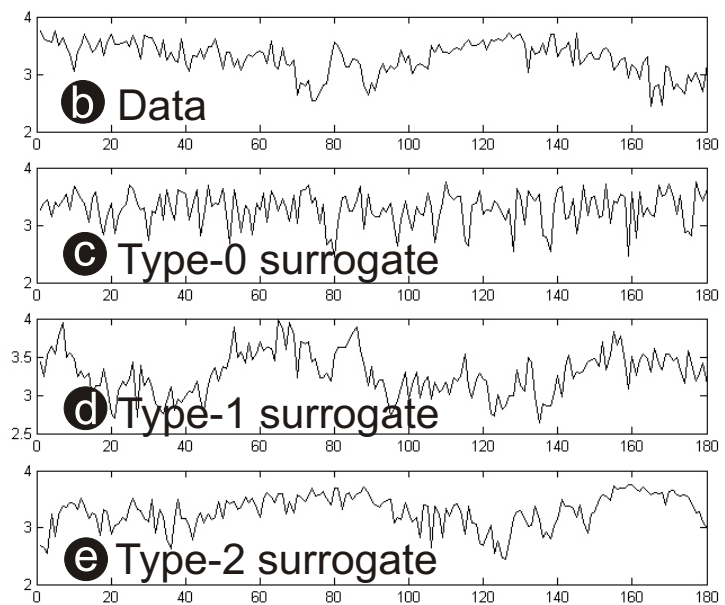
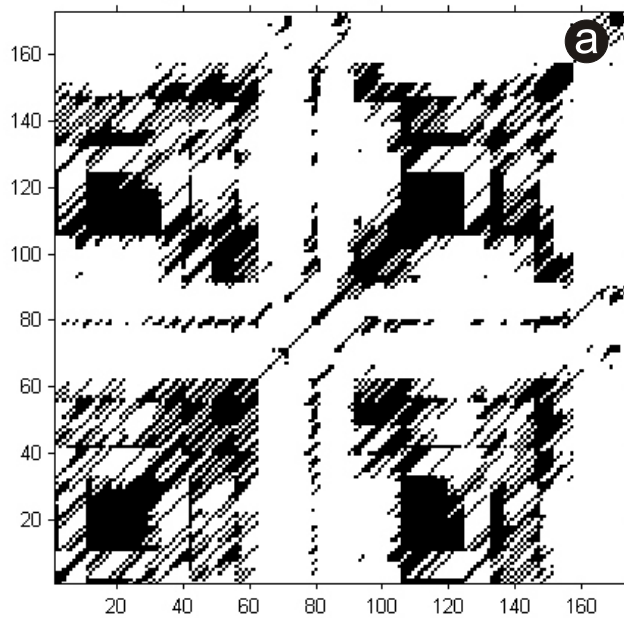


Figure 4

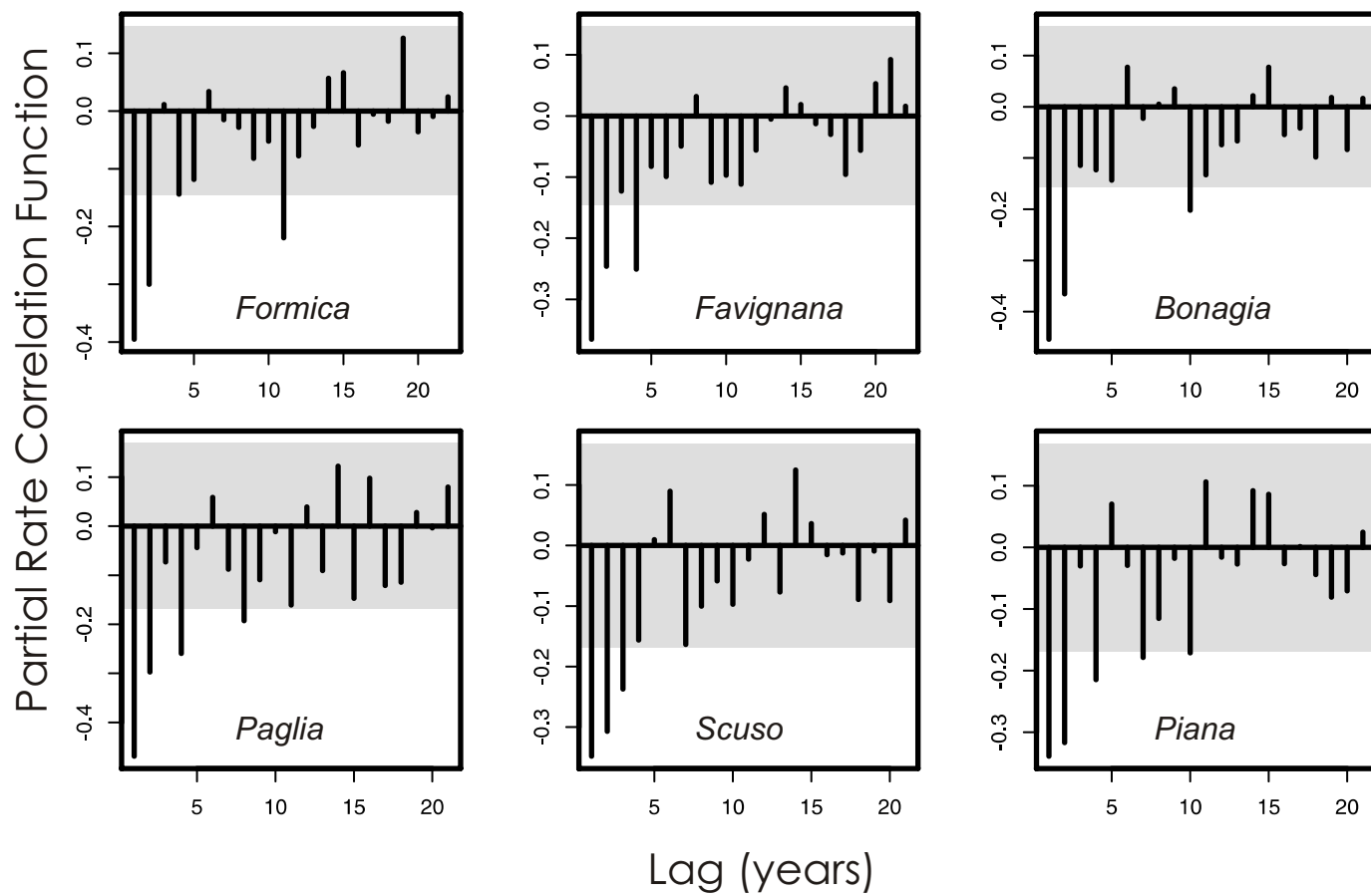


Figure 5

1 **REFERENCES**

- 2 Bakun A, Cury P (1999) The "school trap": a mechanism promoting large-amplitude out-of-
3 phase population oscillations of small pelagic fish species. *Ecology Letters* 2:349-351
- 4 Berryman A (2002) Population regulation, emergent properties, and a requiem for density
5 dependence. *Oikos* 99:600-606
- 6 Berryman A, Turchin P (2001) Identifying the density-dependent structure underlying
7 ecological time series. *Oikos* 92:265-270
- 8 Bjornstad O, Nisbet R, Fromentin J-M (2004) Trends and cohort resonant effects in age-
9 structured populations. *J Anim Ecol* 73:1157-1167
- 10 Boudjema G, Cazelles B (2001) Extraction of nonlinear dynamics from short and noisy time
11 series. *Chaos Soliton Fract* 12:2051-2069
- 12 Boudjema G, Chau NP (1996) Revealing dynamics of ecological systems from natural
13 recordings. *Ecol Model* 91:15-23
- 14 Box GEP, Jenkins GM (1976) *Time series analysis, forecasting and control*, Holden Day, San
15 Francisco
- 16 Cazelles B, Boudjema G (2001) The Moran effect and phase synchronization in complex
17 spatial community dynamics. *Am Nat* 157:670-676
- 18 Chaitin GJ (2001) *Exploring randomness*, Springer-Verlag, London
- 19 Chatfield C (1999) *The analysis of time series - an introduction*, Chapman & Hall
- 20 Cuddington KM, Yodzis P (1999) Black noise and population persistence. *Proc. Roy. Soc.*
21 *London, Ser. B* 266: 969-973
- 22 Dale MRT, Dixon P, Fortin M-J, Legendre P, Myers DE, Rosenberg MS (2002) Conceptual
23 and mathematical relationships among methods for spatial analysis. *Ecography*
24 25:558-577

- 1 Dippner JW, Heerkloss R, Zbilut JP (2002) Recurrence Quantification Analysis as a tool for
2 characterisation of non-linear mesocosm dynamics. *Mar Ecol Prog Ser* 242:29-37
- 3 Eckmann J-P, Kamphorst SO, Ruelle D (1987) Recurrence Plots of Dynamical Systems.
4 *Europhys Lett* 5:973-977
- 5 Farrugio H (1981) Exploitation et dynamique des populations de thon rouge, *Thunnus thynnus*
6 (Linné 1758) Atlanto-Méditerranéennes. PhD dissertation, Université des Sciences et
7 Techniques du Languedoc
- 8 Franks PJS, Wroblewski JS, Flierl GR (1986) Behavior of a simple plankton model with
9 food-level acclimation by herbivores. *Mar Biol* 91:121-129
- 10 Fromentin J-M (2002) Can stochastic variations in recruitment induce long-term fluctuations
11 in the carrying capacity? *Col Vol Sci Pap ICCAT* 54:985-991
- 12 Fromentin J-M, Myers RA, Bjornstad ON, Stenseth NC, Gjosaeter J, Hartvig C (2001) Effects
13 of density-dependent and stochastic processes on the regulation of cod populations.
14 *Ecology* 82:567-579
- 15 Fromentin J-M, Powers JE (2005) Atlantic bluefin tuna: population dynamics, ecology,
16 fisheries and management. *Fish Fish* 6:281-306
- 17 Gao J, Cai H (2000) On the structures and quantification of recurrence plots. *Phys Lett A*
18 270:75-87
- 19 Garcia A, Alemany F, Velez-Belchí P, López Jurado JL, de la Serna JM, González Pola C,
20 Rodriguez JM, Jansá J (2003) Bluefin tuna and associated species spawning grounds
21 in the oceanographic scenario of the Balearic archipelago during June 2001. *Col Vol*
22 *Sci Pap ICCAT* 56:138-148
- 23 Gurney WS, Nisbet R (1998) *Ecological dynamics*, Oxford University Press, Oxford
- 24 Halley JM (1996). *Ecology, evolution, and 1/f noise*. *Trend Ecol. Evol.* 11:33-37
- 25 Hasselmann K (1976) Stochastic climate models I: theory. *Tellus* 28:473-485

- 1 Hilborn R, Mangel M (1997) *The Ecological Detective: Confronting Models with Data*,
2 Princeton University Press, Princeton
- 3 Hsieh C, Glaser SM, Lucas AJ, Sugihara G (2005) Distinguishing random environmental
4 fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature*
5 435:336-340
- 6 Kantz H, Schreiber T (1997) *Nonlinear Time Series Analysis*, Cambridge University Press,
7 Cambridge
- 8 Lioka C, Kani K, Nhhala H (2000) Present status and prospects of technical developmment of
9 tuna sea-farming. *Cahiers Options Méditerranéennes* 47: 275-285
- 10 Lutcavage M, Brill R, Skomal GB, Bradford CC, Howey PW (1999) Results of pop-up
11 satellite tagging of spawning size class fish in the Gulf of Maine: do North Atlantic
12 bluefin tuna spawn in the mid-Atlantic? *Can J Fish Aquat Sci* 56:173-177
- 13 Marwan N (2003) *Encounters With Neighbours - Current Developments Of Concepts Based*
14 *On Recurrence Plots And Their Applications*. PhD thesis, University of Potsdam
- 15 Marwan N, Wessel N, Meyerfeldt U, Schirdewan A, Kurths J (2002) Recurrence Plot Based
16 Measures of Complexity and its Application to Heart Rate Variability Data. *Phys Rev*
17 E 66:026702
- 18 Mather FJ, Mason JM, Jones AC (1995) *Historical document: life history and fisheries of*
19 *Atlantic bluefin tuna*. Report No. 370, NMFS-SEFSC
- 20 Matlab (2003) *Matlab release 13, The language of technical computing*, MathWorks Inc.,
21 Natick, MA, USA
- 22 Medina A, Abascal FC, Megina C, Garcia A (2002) Stereological assessment of the
23 reproductive status of female Atlantic northern bluefin tuna during migration to
24 Mediterranean spawning grounds through the Strait of Gibraltar. *J Fish Biol* 60:203-
25 217

- 1 Morales JM (1999) Viability in a pink environment: why 'white noise' models can be
2 dangerous. *Ecol Lett* 2:228-232
- 3 Parrott L (2004) Analysis of simulated long-term ecosystem dynamics using visual recurrence
4 analysis. *Ecol Complex*:111-125
- 5 Ranta E, Lundberg P, Kaitala V, Laakso J (2000) Visibility of the environmental noise
6 modulating population dynamics. *P Roy Soc Lond B Bio* 267:1851-1856
- 7 Ravier C, Fromentin J-M (2001) Long-term fluctuations in the Eastern Atlantic and
8 Mediterranean bluefin tuna population. *ICES J Mar Sci* 58:1299-1317
- 9 Ravier C, Fromentin J-M (2004) Are the long term fluctuations in Atlantic bluefin tuna
10 population related to environmental changes? *Fish Oceanogr* 13:145-160
- 11 Ripa J, Lundberg P, Kaitala V (1998) A General Theory of Environmental Noise in
12 Ecological Food Webs. *Am Nat* 151:256-263
- 13 Rohani P, Miramontes O, Keeling MJ (2004) The colour of noise in short ecological time
14 series data. *Math Med Biol* 21:63-72
- 15 Royama T (1992) *Analytical population dynamics*, Chapman & Hall, London
- 16 Rudnick DL, Davis RE (2003) Red noise and regime shifts. *Deep-Sea Res* 50:691-699
- 17 Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. *Phys Rev Lett*
18 77:635-638
- 19 Small M, Tse CK (2003) Determinism in Financial Time Series. *Stud Nonlinear Dyn E* 7:1-31
- 20 Spencer PD, Collie JS (1997) Patterns of population variability in marine fish stocks. *Fish*
21 *Oceanogr* 6:188-204
- 22 Steele JH (1985) A comparison of terrestrial and marine ecological systems. *Nature* 313:355-
23 358
- 24 Sugihara G (1994) Non-linear forecasting for the classification of natural time-series. *Philos T*
25 *Roy Soc A* 348:477-495

- 1 Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Young L-S (eds)
2 Dynamical Systems and Turbulence, Vol 898. Springer-Verlag, New York, p 366-381
- 3 Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for non-linearity in
4 time series: the method of surrogate data. *Phys D* 58:77-94
- 5 Tiews K (1978) On the disappearance of bluefin tuna in the North Sea and its ecological
6 implications for herring and mackerel. *Rapp P-v Réun Cons int Explor Mer* 172:301-
7 309
- 8 Turchin P (2003) *Complex Population Dynamics: a Theoretical/Empirical Synthesis*,
9 Princeton
- 10 Vasseur DA, Yodzis P (2004) The color of environmental noise. *Ecology* 85:1146-1152
- 11 Zbilut JP, Thomasson N, Webber Jr. CL (2002) Recurrence quantification analysis as a tool
12 for nonlinear exploration of nonstationary cardiac signals. *Med Eng Phys* 24:53-60
- 13

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

**Environmental Noise in Spawning Areas:
the case of Atlantic Bluefin Tuna**

Proposed running title: “Environmental noise in bluefin tuna spawning areas”

F. ROYER¹ and J.-M. FROMENTIN²

¹Large Pelagics Research Lab, Dept. of Zoology, Univ. of New Hampshire, Durham, NH
03824, USA

²IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean
Monnet, BP 171, 34203 Sète cedex, France

1 **ABSTRACT**

2 Following seminal works on the impact of the colour of environmental noise on
3 population dynamics, we here present an exploratory analysis aiming to seek whether such
4 theory could be original and fruitful to understand the intriguing reproductive strategy of
5 Atlantic bluefin tuna (BFT). A frequency-based approach based on spectral exponents, f^β with
6 $\beta < 0$, is chosen and applied on 10 biogeographical provinces covering the North Atlantic. The
7 major BFT spawning area, i.e. the Mediterranean Sea, was the only one to display a pink
8 power spectrum, whereas open ocean regions displayed more reddened fluctuations, i.e.
9 greater variance at low frequencies. We suggest that environmental noise in the Mediterranean
10 could, thus, offer more favourable characteristics on the long-term than the open ocean. The
11 implications of these findings are discussed in regards to BFT variability in regards to longer
12 (possibly evolutionary) time scales.

13

14 **KEYWORDS**

15 Spawning, environmental noise, spectral exponent, sea surface temperature, *Thunnus thynnus*

16

1 INTRODUCTION

2 The magnitude of environmental stochasticity (i.e. its variance in the time domain) has
3 been long recognized as a key factor to understand population growth and/or extinction in
4 both terrestrial and marine systems (Cushing, 1995, Hjort, 1914, Nisbet & Gurney, 1982).
5 More recently, Lawton (1988) put forward that the colour of environmental forcing (i.e. its
6 variance in the frequency domain) is also a crucial issue. When postulating that reddened
7 noise (i.e. noise with higher variance at low frequencies) could increase population extinction
8 due to a higher probability of consecutive adverse conditions, J.H. Lawton opened an
9 important debate in quantitative ecology. Following studies showed, however, that the
10 mechanisms were more complex than previously thought. The persistence of a population
11 does not solely and directly depend on the colour of environmental noise, but on its
12 interactions with biological processes (Halley, 1996, Morales, 1999, Ripa & Lundberg, 1996,
13 Petchey *et al.*, 1997). For instance, environmental red noise increases population persistence
14 of overcompensatory population dynamics, but decreases this of undercompensatory ones
15 (and inversely for blue noise, Petchey *et al.*, 1997).

16 Marine environmental noise is known to be redder than in terrestrial systems (Steele,
17 1985, Vasseur & Yodzis, 2004). As exploitation reduces population size, fishing could further
18 increase indirectly the risk of extinction of exploited populations through the effects of red
19 noise on depleted populations. The critical situation of the North Sea cod is a good example of
20 this. In the one hand, the long-lasting heavy exploitation has strongly depleted the spawning
21 stock biomass, which is now reduced to a few age-classes of young adults (ICES, 2004). On
22 the other hand, the recruitment success of this stock is negatively related to temperature
23 (Planque & Frédou, 1999, O'brien *et al.*, 2000), so that two decades of warmer conditions (i.e.
24 red noise) increased recruitment failures and thus reinforced the depletion of the stock and
25 extinction risk. Most of the exploited marine populations are particularly prone to these
26 phenomena, since they critically depend on their physical habitat during the fragile but
27 compulsory egg and larvae life stages (Cushing, 1990, Hjort, 1914).

28 Another facet of this issue relates to the spatial distribution of the spawning grounds of
29 large pelagic fish. Comparing life history traits of tuna, Fromentin and Fonteneau (2001) and
30 Schaefer (2001) pointed out a key difference between tropical and temperate tuna. The former
31 display a long spawning period throughout the wide tropical ocean, whereas the latter exhibit
32 a confined spawning window in both space and time. The mechanism of natal homing has
33 been proposed as a possible explanation, through the imprinting of specific environmental
34 factors at early life stages (Cury 1994) and it has been further developed within the

1 metapopulation concept (Fromentin & Powers 2005). One may also ask if the maintenance of
2 such life strategy does not (partially) result from an evolutionary process in relation to the
3 spatial heterogeneity of the environmental forcing. In other words, Atlantic bluefin tuna
4 would only maintain reproduction in confined Mediterranean regions in June (and secondarily
5 in the Gulf of Mexico in May, Fig. 1, see Mather *et al.*, 1995), because the other areas of its
6 wide spatial distribution (i.e. most of the North Atlantic and adjacent seas) would not favour
7 this in respect to their variance in the time and the frequency domain. To examine this
8 hypothesis, we collected the largest *in-situ* sea surface temperature (SST) dataset available.
9 We then computed and compared the seasonal cycles, coefficients of variation and spectral
10 properties of several long term time series of SST of the main oceanographic regions of the
11 North Atlantic and adjacent seas.

1 DATA AND METHOD

2 For this purpose, the *in-situ* measurements of the International Comprehensive Ocean-
3 Atmosphere DataSet (ICOADS, <http://www.cdc.noaa.gov/>) were used. The monthly datasets
4 display a number of geophysical variables at a 2°x2° resolution, starting in 1800. We,
5 however, selected data from the 20th century (1900-2002) to limit spurious sampling
6 variability. For this work, we focused on temperature, a critical factor for growth and survival
7 of eggs and larvae of most fish species, especially for tuna which seldom reproduce under
8 24°C (Schaefer, 2001). For comparison purposes, we considered ten areas based on the
9 biogeographical provinces defined by Longhurst (1998): 1) the Gulf of Mexico and Caribbean
10 Sea (CARB), 2) the western part of the sub-tropical gyre (NASTW), 3) the Gulf Stream
11 region (GFST), 4) the northern Gulf Stream extension (NADR) 5), the eastern part of the sub-
12 tropical gyre (NASTE), 6) the Mediterranean Sea (MEDI), 7) the canarian province (CNRV),
13 8) the northern part of the tropical gyre (NATR), 9) the western tropical province (WTRA),
14 and 10) its eastern counterpart (ETRA, Fig. 1).

15 The number of SST observations was homogeneous between months within each
16 province, with an average of 134 770 samples (46580 std), except for the more restricted
17 Canarian region with 5790 samples on average (470 std). To characterize patterns of SST in
18 the frequency domain, we followed Mandelbrot (1983) and Halley (1996), who showed that a
19 wide range of ‘noises’ in nature can be described by the inverse power law, f^β (where $\beta < 0$ and
20 f is the frequency). β is thus a synthetic measure of the frequency composition of a time
21 series, defined as the slope of the regression of the \log_{10} spectral densities against the \log_{10}
22 frequencies (Vasseur & Yodzis, 2004). Greater the negative slope, higher the degree of
23 autocorrelation in the time series. White noise corresponds to the special case for which $\beta = 0$
24 and can be seen as stochastic variations around a fixed mean. $\beta = -1$ corresponds to red noise,
25 i.e. signals dominated by low frequencies or long-term fluctuations (Vasseur & Yodzis, 2004).
26 Lower β s will yield more to even more “coloured” noises (more autocorrelated time series),
27 i.e. $\beta = -2$ for brownian noise and $\beta = -3$ for black noise. The monthly aggregation induced a
28 cut-off in the frequency spectrum and decreased the effects of weekly and daily variability,
29 but this was not considered critical as BFT spawning usually occurs over 1 to 2 months
30 (Mather *et al.*, 1995). The seasonal component was further removed by subtracting a 12-
31 month anomaly from each series (Chatfield, 1989). A Parzen window was then applied to the
32 raw periodogram to correct for inconsistency, as in Priestley (1981). For each area, this

- 1 procedure was applied to the whole time series as well as to the time series of each given
- 2 month (i.e. January, ..., December).

1 **RESULTS**

2 Considering that BFT (as all tuna species) needs warm waters to reproduce ($> 24^{\circ}\text{C}$,
3 Schaefer, 2001) and that BFT reproduction occurs in May-June, absolute values of
4 temperatures appear limiting in 3 provinces over 10, i.e. NADR, GFST and NASTE (Fig. 2a).
5 The coefficients of variation (CV) of the whole SST time series of the ten oceanographic
6 regions range between 20% (WTRA, CARB and NATR) and 70% (GFST) and mostly reflect
7 the strength of the seasonal variability of the SST (being higher in temperate regions than in
8 the tropical and sub-tropical ones, Fig. 1). Their power spectra are also given in Fig. 1
9 together with the corresponding β values. The β s vary greatly from one area to another, from -
10 0.6 (MEDI) to -1.2 (NATR). SST of the seven main North Atlantic regions (i.e. NASTW,
11 GFST, NADR, NASTE, NATR, WTRA and ETRA) are characterized by typical red noises
12 ($\beta \cong -1$). In these areas, variations in SST are, thus, dominated by low frequencies. The main
13 BFT spawning area (MEDI) displays the flattest SST spectrum, i.e. the variations are more
14 balanced between high and low frequencies (Fig. 1). SST in CARB and CNRY are a bit in
15 between. In other words, MEDI, and secondarily CNRY and CARB, display a larger range of
16 frequencies (especially more short-term variability) than the seven oceanic regions of the
17 North Atlantic (which are more variable on the long-term). Looking at each time series
18 month-by-month reveals another interesting feature (Fig. 2b). NADR, MEDI and CNRY are
19 characterized by even flatter SST spectra during spring (i.e. the BFT spawning period), with
20 β between -0.2 and -0.5, i.e. signals close to white noise with almost as much high as low
21 frequencies. Note that CARB also displays a lower β in May, when BFT spawning occurs in
22 this area.

1 **DISCUSSION**

2 MEDI, which include the major spawning ground for Atlantic bluefin tuna (Ravier &
3 Fromentin, 2001, Mather *et al.*, 1995), displayed the flattest spectrum of all provinces. The
4 SST variability of this area, and secondarily of CNRY and CARB, tends to have a pink
5 spectrum, while the SST fluctuations in the open ocean are more reddened. This complies
6 with the paradigm in physical oceanography postulating that oceanic processes smooth
7 atmospheric white noise (Hasselmann, 1976, Steele, 1985, Steele, 1989). As such, the
8 characteristic time scale (or inertia) of an oceanic basin is related to its size and sensitivity to
9 climate: confined basins are more subject to atmospheric forcing and thus display flatter
10 variance spectra. This is not indication of long-term invariance, as shown by the high CV in
11 the Mediterranean, but rather an indication that a great range of time frequencies is embedded
12 therein. The Mediterranean Sea can thus be perceived as a "miniature ocean", where all the
13 major physical mechanisms occur. A reproductive strategy could there offer to BFT more
14 spatio-temporal options per unit area than oceanic regions. In other words, spawning in short
15 temporal window in enclosed basins could led to higher survival offsprings than in oceanic
16 regions; a feature that is maintained through natal homing (i.e. imprinting of early stages).
17 Note that tropical regions (WTRA and ETRA), where tropical tuna inhabit and spawn over
18 long periods and large areas (Fromentin and Fonteneau, 2001), clearly display redder spectra,
19 a result that is in agreement with the above conclusion.

20 Further, BFT is a long-lived species that includes numerous reproductive age-classes,
21 so that its dynamics are rather resistant to accidents or short-term fluctuations in the
22 recruitment (Fromentin & Fonteneau, 2001). Such life history traits also make that the
23 environmental stochasticity taking place during the BFT reproduction is echoed and
24 distributed within the spawning stock (Bjornstad *et al.*, 2004). Actually, bluefin tuna life cycle
25 acts as a low pass-filter and reddens environmental noise (as opposed to tropical tuna, see
26 Fromentin & Fonteneau, 2001). A feature that is also in agreement with the hypothesis
27 according to which red environmental noise is unfavourable to BFT life cycle.

28 We are aware that this analysis remains too preliminary to give any strong indication
29 on such a complex issue. Our purpose was firstly to draw attention on the fruitfulness, in a
30 fisheries ecology perspective, of an exploratory approach based on the analysis of geophysical
31 datasets. Further work is obviously needed, e.g. to investigate the spawning grounds at finer
32 scales and the use of other environmental factors. While temperature is indeed the most
33 studied geophysical variable, turbulence and retention are also known to be critical for the
34 survival of fish eggs and larvae (Cushing, 1990). Collecting or reconstructing time series of

1 these indices implies tedious work, but could provide new insights on the reproductive
2 strategy of many pelagic species (Agostini, 2000). The colour of environmental noise in
3 marine population dynamics is a fundamental problem, which can be discussed in the light of
4 both life history traits and evolutionary issues.

1 **REFERENCES**

- 2
- 3 Agostini V. (2000) Relationships between small pelagics and the environment: where have we
4 been and where do we go from here? *Report of the 2000 Session of the Working*
5 *Group on small pelagic species*, Fuengirola - Spain
- 6 Bjornstad O., Nisbet R., Fromentin J.-M. (2004) Trends and cohort resonant effects in age-
7 structured populations. *J. Anim. Ecol.* **73**:1157-1167
- 8 Chatfield C. (1989) *The analysis of times series: an introduction*, London:Chapman & Hall,
- 9 Cury P., (1994) Obstinate nature: An ecology of individuals. Thoughts on reproductive
10 behaviour and biodiversity. *Can. J. Fish. Aquat. Sci.* **51**:1664-1673
- 11 Cushing D.H. (1990) Plankton production and year class strength in fish populations: an
12 update of the match/mismatch hypothesis. *Adv. Mar. Biol.* **26**:249-293
- 13 Cushing D.H. (1995) *Population production and regulation in the sea: a fisheries perspective*,
14 Cambridge, Cambridge University Press
- 15 Fromentin J.-M., Fonteneau A. (2001) Fishing effects and life history traits: a case-study
16 comparing tropical *versus* temperate tunas. *Fish. Res.* **53**:133-150
- 17 Fromentin J.-M., J.E. Powers. 2005. Atlantic bluefin tuna: population dynamics, ecology,
18 fisheries and management. *Fish Fish.* (in press)
- 19 Halley J.M. (1996) Ecology, evolution, and 1/f noise. *Trend Ecol. Evol.* **11**:33-37
- 20 Hasselmann K. (1976) Stochastic climate models I: theory. *Tellus* **28**:473-485
- 21 Hjort J. (1914) Fluctuations in the great fisheries of northern Europe. Viewed in the light of
22 biological research. *Rapp. P.-v. Réun. Cons. int. Explor. mer* **20**:1-228
- 23 ICES (2004) Report of the ICES Advisory Committee on Fishery Management and Advisory
24 Committee on Ecosystems, 2004. Vol. 1, Num. 2, Copenhagen

- 1 Lawton J.H. (1988) More time means more variations. *Nature* **334**:563
- 2 Longhurst A. (1998) Ecological Geography of the sea. London, Academic Press
- 3 Mandelbrot B.B. (1983) The fractal geometry of nature. San Francisco, W.H. Freeman
- 4 Mather F.J., Mason Jr J.M., Jones A. (1995) Historical document: life history and fisheries of
5 Atlantic bluefin tuna. Miami, *NOAA Technical Memorandum NMFS-SEFSC-370*.
6 Miami, 165 p.
- 7 Morales J.M. (1999) Viability in a pink environment: why "white noise" models can be
8 dangerous. *Ecol. Lett.* **2**:228-232
- 9 Nisbet R.M., Gurney W.S.C. (1982) Modelling fluctuating populations. Chichester. Wiley
- 10 O'Brien C.M., Fox C.J., Planque B., Casey J. (2000) Climate variability and North Sea cod.
11 *Nature* **404**:142
- 12 Petchey L.G., Gonzalez A., Wilson H.B. (1997) Effects on population persistence: the
13 interaction between environmental noise colour, intraspecific competition and space.
14 *Proc. Roy. Soc. London, Ser.B* **264**:1841-1847
- 15 Planque B., Frédou T. (1999) Temperature and the recruitment of Atlantic cod (*Gadus*
16 *mohrua*). *Can. J. Fish. Aquat. Sci.* **56**:2069-2077
- 17 Priestley M.B. (1981) Spectral analysis and times series. London, Academic Press
- 18 Ravier C., Fromentin J.-M. (2001) Long-term fluctuations in the Eastern Atlantic and
19 Mediterranean bluefin tuna population. *ICES J. Mar. Sci.* **58**:1299-1317
- 20 Ripa J., Lundberg P. (1996) Noise colour and the risk of extinctions. *Proc. Roy. Soc. London,*
21 *Ser.B* **263**:1751-1753
- 22 Schaefer K.M. (2001) Reproductive biology of tunas. In: *Tuna Physiology, ecology, and*
23 *evolution* Block BA, Stevens ED (eds). San Diego, Academic Press. pp 225-270

- 1 Steele J.H. (1985) A comparison of terrestrial and marine ecological systems. *Nature*
- 2 **313**:355-358

- 3 Steele J.H. (1989) The ocean "landscape". *Landscape Ecol.* **3**:185-192

- 4 Vasseur D.A., Yodzis P. (2004) The color of environmental noise. *Ecology* **85**:1146-1152

1 **FIGURES AND TABLES**

2

3 **Figure 1.** Map of the Atlantic bluefin tuna distribution in the North Atlantic and adjacent seas
4 with its known migration routes and spawning areas on which are superimposed the ten
5 biogeographical regions used in this study. The CV, the spectrum (log-log scale) and
6 associated β of the 1900-2002 time series of COADS Sea Surface Temperature are depicted
7 for each province.

8

9 **Figure 2.** Montly SST values (2a) and β coefficients (2b) for the ten biogeographical regions:
10 the Gulf of Mexico and Caribbean Sea (CARB), the western part of the sub-tropical gyre
11 (NASTW), the Gulf Stream area (GFST), the northern Gulf Stream extension (NADR), the
12 eastern part of the sub-tropical gyre (NASTE), the Mediterranean Sea (MEDI), the canarian
13 province (CNRV), the northern part of the tropical gyre (NATR), the western tropical
14 province (WTRA) and its eastern counterpart (ETRA).

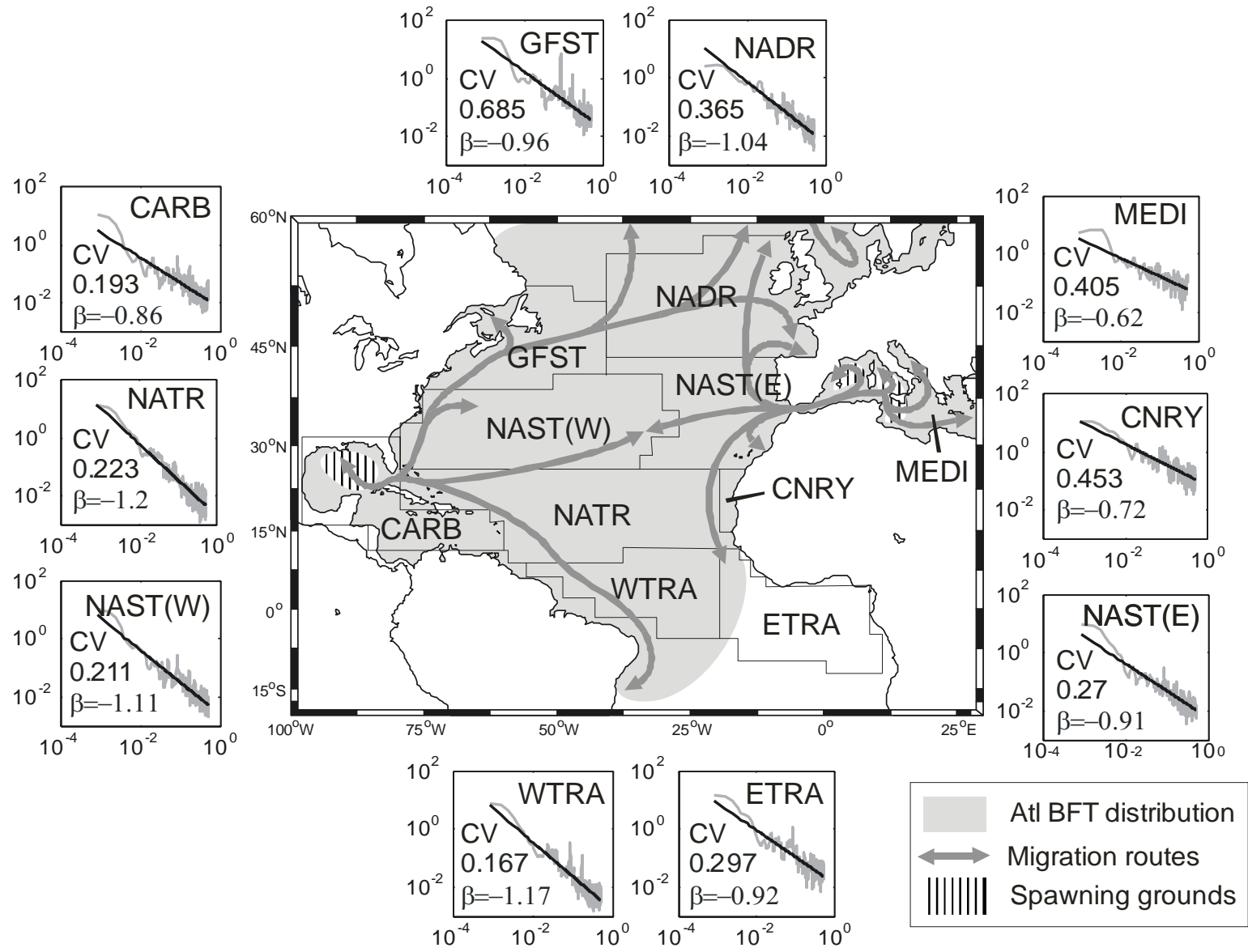


Figure 1

1
2
3
4
5
6

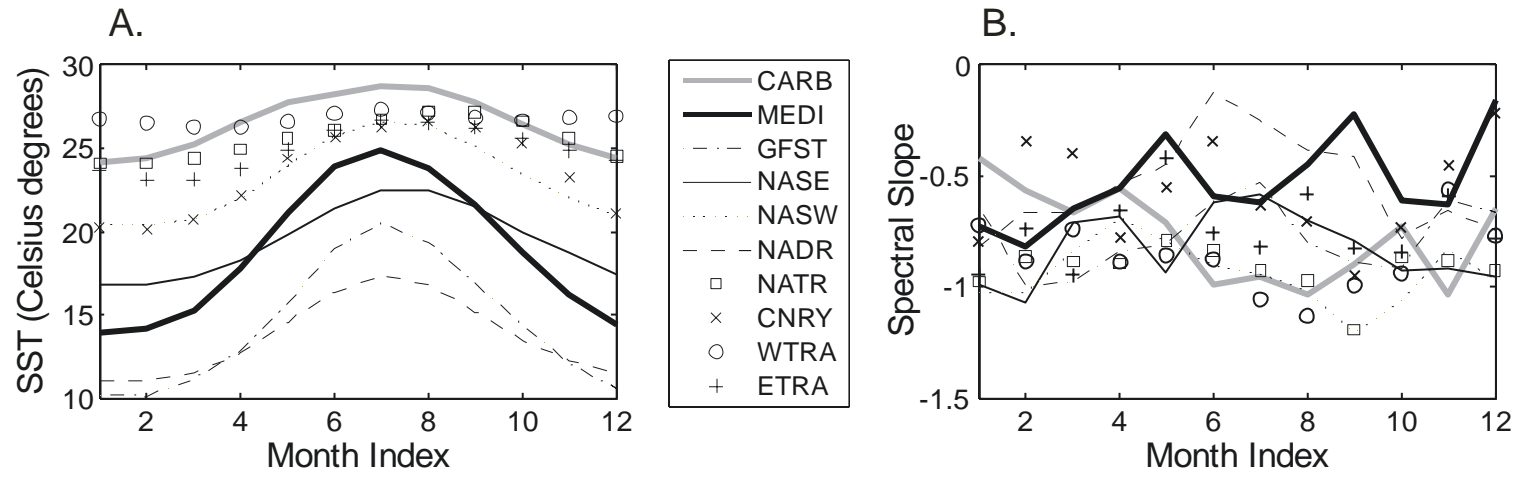


Figure 2

7
8
9
10
11

Abstract

[1] The spatial and temporal dynamics of Atlantic bluefin tuna, *Thunnus thynnus*, are here investigated through data analysis and statistical modelling. The local, regional and global scales are considered successively. A phenomenological approach is favoured, as well the extended use of image processing tools and estimation theory techniques. Their potential role in partially observed or ill-defined problems in marine ecology is presented and discussed. The acute need for frequency-explicit methods is in particular stressed out. Hypothesis testing and significance testing of ecological space/time processes are further discussed.

[2] The first chapter depicts how surface fronts can structure the pelagic habitat: dedicated detection and mapping tools are introduced, along with an example in the North-East Atlantic and the Mediterranean regions. The influence of such heterogeneity on top predators is then discussed and illustrated with a case study in the gulf of Lions: high resolution datasets derived from aerial surveys and satellite radiometry are merged for this purpose. Monte-Carlo methods are shown to be of great value to assess the significance of observed spatial patterns.

[3] The second chapter introduces an individual-based view of the pelagic habitat, through the use of archival tags. Three major drawbacks of this approach are first listed, *i.e.* (i) the large geolocation errors linked to hardware and software issues, (ii) the non-linearity and ill-posedness of the initial estimation problem, and (iii) the lack of accurate and continuous 3D ancillary data sources. A Monte-Carlo method is proposed to address these problems, and applied first to a controlled case. A real world application is then presented, involving pop-up tags and 3D temperature outputs from the Mercator model in the Gulf Stream area. Some specific technical points and issues to resolve are pinpointed.

[4] In the third part of this work, an extended time series analysis is conducted on the long-term fluctuations in the bluefin tuna population, as derived from coastal trap catches in the Mediterranean sea. Linear and non-linear general techniques are employed to uncover deterministic features in these series. A direct and delayed density-dependence is in particular extracted from the high frequency component of the data. Possible endogenous and exogenous causes for the low and high frequency are discussed, as well as the need for new significance tests in studies involving long term changes.

Résumé

[1] L'analyse de données spatio-temporelles est appliquée à l'étude de la dynamique du thon rouge, *Thunnus thynnus*. Les échelles locales, régionales et globales sont considérées successivement dans trois volets différents. L'approche phénoménologique, basée sur la reconnaissance de formes, le filtrage, le lissage et la théorie de l'estimation en général est une caractéristique de ce travail: on y présente les possibilités offertes par ces techniques en écologie marine, et en particulier le besoin d'une décomposition fréquentielle des processus biogéophysiques. Quelques modalités du test d'hypothèse sont discutées, ainsi que le problème de la significativité des résultats en séries temporelles ou dans l'analyse de processus spatialisés.

[2] Le premier volet porte sur l'hétérogénéité spatiale locale de l'habitat pélagique, et des effets des fronts océaniques sur sa structuration. On y présente, illustrés d'un premier exemple, les problèmes de détection et de cartographie des fronts. On aborde ensuite les modalités d'étude des relations fronts/prédateurs pélagiques. Un deuxième exemple basé sur les campagnes de survol STROMBOLI dans le golfe du Lion et des données de radiométrie haute résolution vient illustrer cette partie. On y montre en particulier le besoin de méthodes dédiées (e.g. Monte-Carlo) pour assurer les tests de significativité.

[3] Le deuxième volet aborde l'étude de l'habitat pélagique par l'intermédiaire des marques archives. Trois obstacles majeurs sont identifiés: (i) les larges erreurs de géolocalisation associées à cette technologie; (ii) la non-linéarité et le caractère mal posé du problème, et (iii) le manque de données 3D ancillaires précises associées aux trajectoires. Une méthode de Monte-Carlo est proposée pour filtrer les trajectoire observées, et appliquée à un cas contrôlée pour apprécier ses performances. Une application à un cas réel, basé sur des marques archives de type pop-up et les sorties Mercator dans la région du Gulf Stream, est ensuite décrite et commentée.

[4] On mène dans le troisième volet une analyse des fluctuations à court et long terme du thon rouge, dérivées des séries de capture de madragues méditerranéennes. Des méthodes non-linéaires et linéaires sont combinées pour reconstruire ces fluctuations et mettre en évidence leur structure déterministe. Une densité-dépendance directe et retardée est en particulier extraite du signal à haute fréquence, et mise en valeur par rapport à un éventuel bruit d'observation. Les causes (endogène ou exogène) des pseudo-cycles de grande période sont discutées, de même que les modalités de leur séparation.